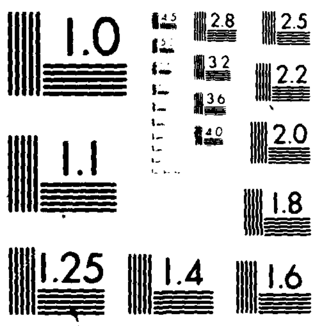
AD-A098 776 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 5/9
A MODEL TO MEASURE BOMBARDIER/NAVIGATOR PERFORMANCE DURING RADA--ETC(U)
MAR 81 T R MIXON

UNCLASSIFIED NL

1 of 4
AD
A098776



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

A MODEL TO MEASURE BOMBARDIER/NAVIGATOR PERFORMANCE DURING RADAR NAVIGATION IN DEVICE 2F114, A-6E WEAPON SYSTEM TRAINER

by

Ted R. Mixon

March 1981

342

Thesis Advisor: W. F. Moroney

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE — READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO. AD-A098 776

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
A Model to Measure Bombardier/Navigator Performance During Radar Navigation in Device 2F114, A-6E Weapon System Trainer

5. TYPE OF REPORT & PERIOD COVERED
Master's Thesis
March 1981

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Ted R. Mixon

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

12. REPORT DATE
March 1981

13. NUMBER OF PAGES
301

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

A-6	Simulator Training
A-6 Flight Training	Simulator Performance
A-6E Weapon System Trainer	Bombardier/Navigator Performance
Aircraft Simulator	Human Performance
Training Devices	Performance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis modeled a performance measurement system for the Bombardier/Navigator (B/N) Fleet Replacement Squadron student during low level radar navigation flight in the Navy A-6E Weapon System Trainer. The model was designed to determine student skill acquisition measures for the purpose of providing information for decision-making by the squadron instructor and training manager. Model formulation methodology

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (continued)

was based on a literature review of aircrew performance measurement from 1962-1980 and an analytical task analysis of the B/N's duties. Over 50 currently accessible candidate measures were listed and a proposal was made for a competitive exercise ~~(Derby)~~ using A-6E fleet aircrews flying preprogrammed routes to establish performance standards using the candidate measures. Multivariate discriminate analysis was recommended for measure reduction. A sequential sampling decision model selected for evaluation provided fixed decisional error rates for successful training termination decisions and ~~utilized~~ used both objective and subjective performance measures. Several display formats were recommended for use in debriefing.

A (used)

Block 19 (continued)

Performance measurement
Performance criteria
Performance tests
Measurement
Pilot performance
Pilot performance measures
Pilot performance measurement
Aircrew performance
Aircrew performance measurement
Objective measurement
Student measurement
Student achievement
Training evaluation
Evaluation
Sequential testing
Quantitative methods
Proficiency
Pilot proficiency
Instructor pilot
Task analysis
System analysis
Time Line Analysis
Individual differences
Skill
Flight skills
Navigation skills
Complex skill acquisition
Navigator

Approved for public release; distribution unlimited

A Model to Measure Bombardier/Navigator Performance During Radar Navigation in Device 2F114, A-6E Weapon System trainer

by

Ted R. Mixon
Lieutenant, United States Navy
B.S., United States Naval Academy, 1974

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1981

Author: Ted R. Mixon

Approved by: William F. Moroney
Thesis Advisor

Second Reader

Chairman, Department of Operations Research

Dean of Information and Policy Sciences

ABSTRACT

This thesis modeled a performance measurement system for the Bombardier/Navigator (B/N) Fleet Replacement Squadron student during low level radar navigation flight in the Navy A-6E Weapon System Trainer. The model was designed to determine student skill acquisition measures for the purpose of providing information for decision-making by the squadron instructor and training manager. Model formulation methodology was based on a literature review of aircrew performance measurement from 1962-1980 and an analytical task analysis of the B/N's duties. Over 50 currently accessible candidate measures were listed and a proposal was made for a competitive exercise (Derby) using A-6E fleet aircrews flying preprogrammed routes to establish performance standards using the candidate measures. Multivariate discriminate analysis was recommended for measure reduction. A sequential sampling decision model selected for evaluation provided fixed decisional error rates for successful training termination decisions and utilized both objective and subjective performance measures. Several display formats were recommended for use in debriefing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Since 1929, when Edwin A. Link produced the first U.S.-built synthetic trainer designed to teach people how to fly, flight simulation has witnessed substantial advances in simulation technology and increased incorporation of these devices into both military and civilian aviation training programs. A recent addition in 1980 to this advance was the A-6E Weapon System Trainer (WST), device 2F114. A sophisticated flight simulator with a six degree of freedom motion system, the A-6E WST was designed to provide the capability for pilot transition training, Bombardier/Navigator (B/N) transition training, integrated crew training, and maintenance of flight and weapon system proficiency in all non-visual elements of the A-6E Carrier Aircraft Inertial Navigation System (CAINS) mission. The development of high-fidelity flight simulation has been accompanied by advances in aircrew performance measurement systems, which are ideal for the simulator training environment, and have been widely implemented and the subject of extensive research in all three military aviation communities.

The purpose of this thesis is to design a system to improve current performance measurement techniques for the B/N Fleet Replacement Squadron (FRS) student by the development and application of a performance measurement system that incorporates the advantages of both objective skill acquisition measures and subjective instructor measurement.

A. BACKGROUND

While simulation has been a popular component of many aviation training systems for over forty years, objective performance assessment had not been incorporated until some fifteen years ago. This section will discuss the current state of flight simulation, aircrew performance measurement, and provide a brief review of previous navigator performance measurement studies.

1. Simulation and Performance Measurement

Simulation is the technique of reproducing or imitating some system operation in a highly-controlled environment. Modern flight simulators have evolved from simple procedure trainers into devices that represent specific aircraft counterparts, and imitate or duplicate on-board systems and environmental factors. The two main purposes of flight simulators within the training environment are training and evaluation. Training is designed to improve performance and some means of providing feedback to the student is needed to indicate the adequacy of his behavior, and ought to provide guidance for the correction of inappropriate response patterns. Evaluation involves testing and recording the student's behavior in the performance examination situation. [Angell, et al., 1964].

The reasons for using the simulator as an integral part of a military flight training program were examined by

Tindle [1979], Shelnutt, et al. [1980], North and Griffin [1977], and Roscoe [1976]. Some basic justifications for simulator training include:

(1) Simulation provides training in skill areas not adaptable to an actual training flight because of technological or safety considerations.

(2) Crews master skills in the aircraft in less time after learning those skills in a simulator.

(3) The cultivation of decision-making skills is an instructional objective calling for situational training that may be carried out safely only in a simulated tactical environment.

(4) Simulators are effective for training crewmembers of varying experience and expertise in a variety of aircraft for a number of flight tasks.

(5) Greater objectivity is obtainable for measuring student performance by using controlled conditions and automated performance measurement features in the simulator than in the aircraft.

(6) Instructors are not distracted by operational constraints in the simulator and are more available for teaching and evaluation roles.

These considerations are by no means exhaustive, but they do indicate the utility of simulators in flight training programs, especially in evaluating student performance.

This thesis is addressed primarily to the problem of measuring B/N performance during a radar navigation training flight while in the A-6E WST. The performance of a B/N is the exerted effort (physical or mental) combined with internal ability to accomplish the A-6E mission and its functions. Some development of performance measurement definitions and goals is necessary because the problem of assessing B/N performance during a training program is obviously but a segment of a broader topic - the measurement of human behavior.

Glaser and Klaus [1966] defined performance evaluation as the assessment of criterion behavior, or the determination of the characteristics of present performance or output in terms of specified standards. The importance of defining performance evaluation is paramount to any training assessment situation, as it gives common ground to operationally describing the human behaviors that make up performance itself, and identifies behavior elements that may be measured by either objective or subjective means. An expanded discussion of performance measurement and evaluation can be found in Chapter IV.

The purposes for assessing performance by the application of standard objective measurement operations were stated by Angell, et al. [1964], and Riis [1966]:

(1) Achievement - to determine the adequacy with which an activity can be performed at the present time, without regard, necessarily, for antecedent events or circumstances.

(2) Aptitude - to predict the level of proficiency at which a person might perform some activity in the future if he were given instructions concerning the activity.

(3) Treatment efficacy - to observe the effects upon performance of variation in some independent circumstances such as (a) instructional techniques, (b) curriculum content, (c) selection standards, (d) equipment configurations, or the like.

The flight simulator training environment allows for special applications of the above as found in Danneskiold [1955], Angell, et al. [1964], Riis [1966], Glaser and Klaus [1966], Farrell [1974], Shipley [1976], and McDowell [1978]:

(1) Diagnostic - determine strong and weak areas of student proficiency.

(2) Readiness - determine operational readiness of an aviation unit.

(3) Discrimination - assess performance to provide information about an individual's present behavior as compared to other individuals.

(4) Selection - of persons for promotion or advancement or placement.

(5) Learning rates - determining the rate at which learning takes place.

(6) Management - of an entire training program and its subsystems.

(7) Evaluation - of training devices in terms of effectiveness and transfer-of-training.

The above goals of performance measurement represent some of the major reasons why assessment of student performance is important in training program simulators. Most importantly, measurement provides FRS instructors and training officers with the information needed to make correct decisions [Obermayer, et al., 1974; Vreuls and Wooldridge, 1977]. Performance measurement does not in itself replace the decision-maker in the FRS, but instead provides complete and necessary information of an objective nature to the appropriate evaluator (instructor or training officer), so that more accurate and reliable decisions can be made concerning student progress within the training syllabus. If instructors and training officers utilize the potential of a performance measurement system, a more effective and efficient training program would be a result.

2. Review of Previous Studies

Most of the literature on aircrew performance measurement in the last forty years has primarily concentrated on the pilot crewmember [Ericksen, 1952; Danneskiold, 1955; Smode, et al., 1962; Buckhout, 1962; Obermayer and Muckler, 1964; Mixon and Moroney, 1981]. The first comprehensive evaluation of techniques used in flight grading was by Johnson and Boots [1943], who analyzed ratings given by instructors and inspectors to students on various maneuvers throughout stages of training. One result showed correlations between grades assigned by different raters to the same subject as being very low. This result of low observer-observer reliability when using subjective ratings will be discussed in the next section.

The earliest studies involving the radar navigation performance of a crewmember other than the pilot were a pen and pencil radar scope interpretation experiment by Beverly [1952], and two Air Force radar bombing error projects by Voiers [1954], and Daniel and Eason [1954]. The first study was concerned with constructing a suitable test for the measurement of navigational radar scope interpretation ability of student aircraft observers. The latter two studies were concerned with identifying perceptual factors which contributed to cross-hair error during bomb runs of a radar bombing mission and with comparing the components of simulated radar bombing error in terms of reliability and sensitivity to practice, respectively. A similar follow-up study on radar scope interpretation and operator performance in finding and identifying targets using a radar was performed by Williams, et al. [1960]. These four studies represent most of the research of non-pilot radar navigation performance measurement prior to 1965. This fact is not surprising, due mainly to the early role played by the observer in very simplified and pilot-oriented aircraft as compared to today's specialized navigator in complex, computer-oriented aircraft.

Since navigation is a primary duty of any aviator across a spectrum of aircraft types, some helicopter pilot and copilot studies are of some value to review. Helicopter Nap-of-the-Earth (NOE) flight is a visual-dominated low level mission where altitude and airspeed are variable in close proximity to the ground. Some navigational performance measures utilized in these studies were: number of turn points found, probability of finding a turn point, and route excursions beyond a criterion distance [Fineberg, 1974; Farrell and Fineberg, 1976; Fineberg, et al., 1978; Smith, 1980]. Low level visual navigation flights in helicopters were also studied in some detail, again with pilot performance being the main concern [Lewis, 1966; Billings, et al., 1968; Sanders, et al., 1979].

Two rather novel investigations of Anti-Submarine Warfare (ASW) helicopter team performance using the content and flow of team communications during simulated attacks were done by Federman and Siegel [1965] and Siegel and Federman [1968].

All team communications were recorded, classified, and compared to target miss distance as an effectiveness measure. Although they found some types of team communication to correlate highly with mission success, their method was highly impractical for the operational situation due to the large number of personnel needed to play back and classify the communication types. Nevertheless, the results from this research indicate the value of using crew communication as a measure of crew performance.

Several fixed-wing studies with navigation as the primary mission are also of interest to the current study. Schohan, et al. [1965] and Soliday [1970] used the Dynamic Flight Simulator for several Low Altitude High Speed (LAHS) missions designed to investigate pilot and observer performance during turbulent, lengthy (3-hour) flights. Jensen, et al. [1972] did several studies investigating pilotage errors in area navigation missions for the Federal Aviation Administration. These three studies are significant in that numerous navigational accuracy performance measures were used to assess pilot (or observer) performance.

After 1970, due to the increased complexity of many modern aircraft, more research was directed toward individual aircrew members, and not just the pilot. Among the aircraft investigated were: P-3C, F-4J, A-7, F-106, B-52, C-141, C-130, KC-135, and the C-5 [Matheny, et al., 1970; Vreuls and Obermayer, 1971; Obermayer and Vreuls, 1974; Geiselhart, et al., 1976; Swink, et al., 1978]. These studies are unique in that defining and assessing aircrew performance by other than the subjective ratings, as commonly used for decades, became a technological challenge requiring new analytical and empirical approaches.

An Air Force fighter-bomber, the F-111D, was designed and built during the late 1960's with virtually the same tactical capability of the A-6E. With a two-man side-by-side cockpit arrangement, this land-based aircraft is the closest counterpart to the A-6E for the radar navigation air interdiction mission. Two experiments using the F-111A flight simulator were performed mainly for equipment configuration effects on pilot performance [Geiselhart, et al., 1970; Geiselhart, et al., 1971]. Research by Jones [1976] examined the use of the F-111D flight simulator as an aircrew performance evaluation device. Unfortunately, these F-111 studies do not specifically address the issue of how to measure navigator performance during radar navigation, but do provide some information on measuring performance in an aircraft with a similar mission and almost the same crew interactions as the A-6E.

Only one experiment known to this author has been conducted using an A-6 configured simulation. Klier and Gage [1970] investigated the effect of different simulation motion conditions on pilots flying air-to-air gunnery tracking tasks in the Grumman Research Vehicle Motion Simulator (RVMS).

They concluded that simulator motion need not be a faithful reproduction of real-life motion in order to provide essential motion cues. Saleh, et al. [1980] performed an analytical study for two typical tactical combat missions representative of the A-6E and A-7E aircraft to determine significant decisions which are made in the course of accomplishing mission objectives. The results of this study provide information regarding the decision type, difficulty, and criticality and can be used in identifying the critical areas in which aircrew decision-aiding may significantly improve performance. Finally, a study by Tindle [1979] concluded that the integration of the A-6E WST (device 2F114) into the FRS training program would be more cost-effective than using the A-6E WST as an addition to existing training programs. This study also concluded that aircrew performance measurement in the A-6E WST was vital for more effective use of the simulator.

This section has presented a brief review of aircrew performance measurement studies in the literature. The potential value in reviewing the literature lies in uncovering the analytical and empirical approaches taken in measuring aircrew performance, noting both the significance and practicality of those approaches. Since previous research on actual B/N performance during the radar navigation air interdiction mission appears to be nonexistent, extrapolations from other aircrew performance measurement studies is important and necessary to the current study. What has worked and been practical for other aircraft and aircrews in the way of performance measurement certainly applies to the A-6E B/N, keeping in mind the definitions and goals of the A-6E B/N, performance measurement, and the A-6E mission.

B. SUBJECTIVE AND OBJECTIVE PERFORMANCE MEASUREMENT

1. Introduction

Traditionally, all aircrew performance measurement in the Navy, Air Force and Army has been assessed by an instructor pilot or navigator using a subjective rating scale which places the student in one of several skill categories based on norm-referenced testing. More recently, objective methods of evaluating performance have been developed and implemented in both the simulator and in-flight environments. Subjective and objective methods are not dichotomous but represent a continuum of performance measurement. At one extreme there exists the strictly personal judgement and rating of performance, and on the other end of the continuum is a completely automated performance measurement and assessment system.

This section will define and describe the elements of each method together with the advantages and disadvantages associated with each. The approach taken in this study will be to integrate the use of automatic performance measurement within the A-6E training environment while still exploiting the advantages of using the instructor as a component of the measuring system.

2. Subjective Performance Measurement

Subjective measurement can be defined as an observer's interpretation or judgement between the act observed and the record of its excellence with an almost complete reliance placed on the judgement and experience of the evaluator [Cureton, 1951; Ericksen, 1952]. Simply stated, subjective measurement is qualitative in nature as what is being measured is observed privately [Danneskiold, 1955; Knoop and Welde, 1973; Roscoe, 1976; Vreuls and Wooldridge, 1977]. Through an introspective process, the "expert" instructor judges the performance level demonstrated by a student whether or not agreed-upon standards of performance have been applied [Billings, 1968; McDowell, 1978].

a. Advantages of Subjective Measurement

The advantages of using subjective performance measurement methods have been well-documented throughout the literature. Instructor ratings in the past have been the least expensive of all evaluation methods [Marks, 1961]. This decisive advantage has been eroded in recent times by severe shortages of military aircrew in both operational and training units. Marks [1961] also pointed out that ratings forms are constructed easily and quickly, and the administration of the ratings system requires no physical arrangement. McDowell [1978] recently concluded that a subjective performance measurement system for many complex tasks such as flying were easy to develop, gave the rating instructor high face validity since he is usually an acknowledged expert, and contained specific feedback of a type important in the training situation usually not found in objective performance measurement systems. Ratings are still used because they meet the needs of training management without seriously intruding into the instructor pilot's operational capabilities [Shipley, 1976]. One important use of an instructor to subjectively grade a student is to motivate the student through selective reinforcement [Prophet, 1972; Carter, 1977]. Sometimes an overly positive or negative grade by the instructor in the appropriate area of desired performance improvement for the student serves as a catalyst in the student's attitude toward self-improvement.

Some studies have shown that some degree of high reliability can be achieved between instructor ratings [Greer, et al., 1963; Marks, 1961]. Brictson [1971], in a study of over 2500 carrier arrested landings, reported measures derived from the Landing Signal Officer (LSO) grades to be highly correlated with objective estimates derived from a weighted combination of wave-offs, bolters, and the particular wire engaged. Similar high correlations between raters were also found by Waag, et al. [1975] for undergraduate pilots flying seven basic instrument maneuvers in the Air Force Advanced Simulation in Undergraduate Pilot Training (ASUPT) facility.

b. Disadvantages of Subjective Measurement

High reliability between instructors using subjective rating methods is generally not the case and therein lies the foremost disadvantage of the subjective performance measurement method. Knoop [1973] reported two instructor pilots (IP's) subjective ratings were each correlated with certain objective performance measures, but the objective measures themselves were determined to be not highly correlated with skilled or proficient operator performance. Ericksen [1952] reviewed numerous flight studies involving pilot training between 1932 and 1952 and concluded that subjective grading involved a lack of reliability and inconsistent differentiation between students. Danneskiold [1955] found observer-observer correlations no higher than .47 for three Basic Instrument Check maneuvers, while a more objective test had observer-observer correlations of .86.

The training and evaluation skills of the instructor evolve primarily from their personal experiences in the highly complex aircraft and simulator environment. Establishing adequate standards of performance, or criteria, is a major problem in all flight training. Knoop and Welde [1973] found lack of agreement between pilots on the specific criteria for successful performance of certain aerobatic maneuvers, due largely to the differences in examiner knowledge, experience, and proficiency. It was also found that the same maneuver may be flown satisfactorily in a number of different ways. Other research has explored the criteria problem which is inherently part of subjective performance measurement [Cureton, 1951; Danneskiold, 1955; Marks, 1961; McDowell, 1978]. Even rating methods were inadequate when used as criteria to validate alternative methods of measurement and evaluation [Knoop and Welde, 1973].

An instructor must be able to process large quantities of information during a simulator session. Subjective grading competes with this capability of the instructor, and may prevent perception and evaluation of all the relevant dimensions of task performance during a training mission [Knoop, 1968; Roscoe, 1976; Shipley, 1976; Carter, 1977; Vreuls and Wooldridge, 1977].

Several other factors which contribute to subjective aircrew rating variances are discussed below:

(1) A tendency of raters to be more lenient in evaluating those whom they know well or are particularly interested in [Smode, et al., 1962; Bowen, et al., 1966].

(2) The observations tend to accumulate on one or two points in the rating scale, usually near the central or average point. This phenomenon contributes toward a lack of sufficient discrimination among students [Smode, et al., 1962; Shipley, 1976].

(3) Instructor and student personalities interact to yield a result which does not reflect true performance [Marks, 1961].

(4) A tendency for the ratings on specific dimensions to be influenced by the rater's overall impression of the student's performance - the "halo effect" [Smode, et al., 1962; Glaser and Klaus, 1966].

(5) An unrelated problem with the simulator may influence the evaluation outcome [Jones, 1976].

(6) Evaluation results are dependent upon the attitude, concern, and values of the instructor, thus a natural personal bias is introduced into the performance observation [Smode, et al., 1962; Knoop and Welde, 1973; Jones, 1976].

(7) Instructors have different concepts of the specific grading system in regard to the flight parameters involved, knowledge tested, weights to be assigned, and ranges of qualifying categories [Knoop and Welde, 1973].

(8) A tendency to actually rate others in the opposite direction from how the rater perceives himself on the particular performance dimension has been found [Smode, et al., 1962].

(9) Ratings tend to become more related along different dimensions when they are made closer to each other in time than ratings having a larger interval of time between observations [Smode, et al., 1962].

(10) Unless the simulator has a playback capability, a permanent record of the performance is lost when subjective ratings are used [Forrest, 1970; Gerlach, 1975].

3. <u>Objective Performance Measurement</u>

Objective measurement is defined as observations where the observer is not required to interpret or judge, but only to record his observations [Cureton, 1951]. While subjective judgement is more qualitative in nature, objective measurement demands that what is being measured be observed publicly and with a quantitative result [Knoop and Welde, 1973; Roscoe, 1976; Vreuls and Wooldridge, 1977; McDowell, 1978]. Objective measures demand that performance be evaluated in terms of criteria which are relatively independent of the observer, have consistent interpretations, and a high degree of observer-observer reliability [Ericksen, 1952; Danneskiold, 1955; Marks, 1961; Smode, et al., 1962].

The first systematic use of an objective grading method was by Miller, et al. [1947]. Objective measures were collected during a single week from over 8,000 students in four different phases of pilot training. Objective observations by way of a prepared checklist reduced variability attributable to the observer and correlations as high as .88 were found between instructors observing a student during the same flight. In most cases, higher observer-observer reliability has been found when objective measures are used [Angell, et al., 1964; Forrest, 1970]. The measures are free from personal and emotional bias of the instructor, as well as judgemental bias that are characteristic of subjective measurement.

a. Advantages of Objective Measurement

Most advantages of objective performance measurement appear to contrast the disadvantages of subjective measurement. By having a computer process large amounts of continuously varying information, the instructor is freed to concentrate on those aspects of student performance which resist objective measurement, and to devote more attention to the primary duty of instructing [Krendel and Bloom, 1963; Knoop and Welde, 1973; Vreuls and Obermayer, 1971, Vreuls, et al., 1974]. Development of performance criteria could be made on the basis of permanently recorded objective measures [Forrest, 1970; Angell, et al., 1964]. A system of data collection would provide records and transcriptions of individual and crew performance in practice missions to identify particularly effective or ineffective behaviors for later analysis in the event of an aircraft accident [Forrest, 1970; Angell, et al., 1964].

Objective measurement enables timely and diagnostic information of consistent weaknesses in performance [Angell, et al., 1964; Knoop and Welde, 1973]. Instructional methods may be modified as performance results indicate their effectiveness, or lack of it. Students attaining desired achievement levels may also be identified earlier within the training syllabus. Several researchers postulate the quantification of skill learning rates [Angell, et al., 1964; Knoop and Welde, 1973]. Bowen, et al. [1966], even found objective measures motivated pilots to actualize their skills in overt performance measurement. They speculated that the "heightening of performance is due to intrinsic motivation (personal desire to achieve), social motivation (pressures from group to demonstrate proficiency), and a focusing of attention on each particular performance which encourages the pilot to actualize the knowledge and skills which he possesses."

b. Disadvantages of Objective Measurement

If objective performance measures are so much more desirable than subjective ratings, why have they not been incorporated into more training situations? The major reason lies in the fact that they are much more expensive than subjective methods and some aircrew tasks are difficult to automatically record and computer grade. Whenever a number of simultaneously occurring tasks such as communication, procedures, and the application of knowledge are present during a particular task, measuring and quantifying the operator behavior involved becomes a complex and inherently difficult task in itself. Smode [1966] found that when flight instructors are required to monitor and record performance information during a flight task, some resentment against the objective measurement method occurred due to the large amount of instructor attention required. This instructor monitoring and recording of student performance information has since been replaced by automatic digital computers. The same report also concluded that detailed analyses of aircrew tasks into perceptual-motor tasks, procedural tasks, and decision-making distorts reality as all three are very interrelated. Objective tests are rigidly constrained by given hardware that has to be physically arranged, programmed, and is subject to equipment malfunctions. Since objective tests require more structure than subjective testing, instructors are given very little choice in what they must do and "there is a certain natural resentment against the regimentation of setting up and observing this event at this time [Smode, 1966]."

4. Summary

Subjective and objective performance measurement of aircrew has been defined, compared, and contrasted for strengths and weaknesses. Subjective testing is universally feasible, minimizes paperwork, allows for instructor flexibility, is easy to develop and administer, and is inexpensive. Objective testing minimizes instructor bias, eases grading due to automatic data collection, storage and dissemination, improves performance standardization, and produces a high inter-rater reliability. Each method has its merits individually, but when used together in a cohesive and synergistic combination, improvements can be made in aircrew performance measurement and assessment.

Angell, et al. [1964] stated, "There are some areas in which the human observer can make more subtle judgements and more sophisticated evaluations than can any electromechanical instruments. . . the human observer/teacher should not be an adjunct, but rather an integral part of the total measurement system." This report will attempt to use this observation in the design of a system to measure A-6E B/N performance during radar navigation in the simulator. This section is concluded by a listing of items the examiner should evaluate, as determined by Knoop and Welde [1973]:

(1) Ability to plan effectively.

(2) Decision-making capability.

(3) Sensorimotor coordination and smoothness of control.

(4) Ability to share attentions and efforts appropriately in an environment of simultaneous activities.

(5) Knowledge and systematic performance of tasks.

(6) Confidence proportionate to the individual's level of competence.

(7) Maturity; willingness to accept responsibility, the ability to accomplish stated objectives, judgements, and reaction to stress, unexpected conditions, and aircraft emergencies.

(8) Motivation (attitude) in terms of the manner in which it affects performance.

(9) Crew coordination.

(10) Fear of flying.

(11) Motion sickness.

(12) Air discipline - adherence to rules, regulations, assigned tasks, and command authority.

C. A-6E TRAM AIRCRAFT AND ITS MISSION

1. A-6E Performance Specifications

The A-6E aircraft is a two-man, subsonic, twin engine medium attack jet aircraft, with side-by-side seating for the pilot and B/N. Designed as a true all-weather attack aircraft using a sophisticated radar navigation and attack system, the aircraft can accurately deliver a wide variety of weapons without the crew ever having visually acquired the ground or the target. Capable of carrying a payload of up to 8.5 tons, it is the only carrier-based aircraft in the Tactical Air

(TACAIR) wing capable of penetrating enemy defenses at night, or in adverse weather, to detect, identify and attack fixed or moving targets. The TRAM (Target Recognition, Attack, Multisensor) configured A-6E aircraft has a completely integrated computer navigation and control system, radar, armament, flight sensors, and cockpit displays that enable the aircraft to penetrate enemy defenses at distances approaching 600 nautical miles in radius while at an extremely low altitude.

2. Mission

The mission of the A-6 "Intruder" is to perform high and low altitude all-weather attacks to inflict damage on the enemy in a combat situation. TACAIR recognizes three primary missions to accomplish the objective of successfully waging war [Gomer, 1979]. The missions are: Close Air Support (CAS), Counter Air (CA), and Air Interdiction (AI). CAS is air action against hostile ground targets that are in close proximity to friendly ground forces, requiring detailed integration of each air mission with the battle activities and movements of those forces. CA operations involve both offensive and defensive air actions conducted to attain or maintain a desired degree of air superiority by the destruction or neutralization of enemy air forces. AI missions are conducted to destroy, neutralize, or delay the enemy's military potential before it can be brought to bear against friendly forces, usually at far distances not requiring detailed integration of air and ground activities. The AI mission was selected in this study because it is representative of those missions frequently performed in the A-6 community. Analysis of all three primary TACAIR missions was beyond the scope of this study.

Saleh [1980] defined a mission as the aggregate scenarios, maneuvers, and segments that constitute successful employment of the system. The starting point in determining criteria for performance measurement and suggesting what specific and clearly identifiable operations of the B/N should be examined in greatest detail is an operational definition of the man-machine mission [Smode, 1962; Vreuls, 1974]. McCoy [1963] further stated that in order to judge the effectiveness of any element of a man-machine system, it must be judged in terms of contribution of the element to the final system output, which is the ultimate objective of the man-machine system. It is with these criteria in mind that the AI mission definition is used to limit performance measurement of the B/N in the A-6E WST.

3. Scenarios

Analysis for comprehensive performance measurement begins with a complete decomposition of the mission into smaller parts for which activities and performance criteria are more easily defined [Vreuls, 1974; Connelly, 1974; Vreuls and Cotton, 1980]. Any mission may be described in terms of a scenario, or intended flight profile or regime. A performance measurement standards tri-service project [Vreuls and Cotton, 1980], classified military aviation into ten scenarios or flight regimes: (1) Transition/Familiarization, (2) Navigation, (3) Formation, (4) Instruments, (5) Basic Fighter Maneuvers, (6) Air Combat Maneuvering, (7) Air-to-Air Intercept, (8) Ground Attack, (9) Air Refueling, and (10) Air Drop. Scenarios may be further subdivided into maneuvers by identifying natural breakpoints using time, position, or definitive portions requiring computation of different performance measures or changes in required operator skill level. Examples of maneuvers are take-off, climb, landing, and point-to-point navigation. Segments are subdivisions of maneuvers that contain groupings of those activities that must be accomplished in performing the maneuver. Table I (adapted from Vreuls and Cotton [1980]) contains possible maneuvers and segments for the navigation scenario.

The present study selected point-to-point navigation using radar terrain mapping to further narrow the scope of the effort and to tailor the performance measurement aspects of the A-6E mission toward the tasks of the B/N.

4. Summary

The mission of the A-6 for the current study has been defined as Air Interdiction which is further subdivided into point-to-point navigation using radar terrain mapping. Successful accomplishment of the radar navigation point-to-point segments reasonably infers some degree of overall Air Interdiction mission accomplishment, which is the overall objective of the A-6 man-machine system. Operationally dividing the overall A-6 mission into flight maneuvers enables practical performance measurement as the operator skills required (and measured) vary from segment to segment. Within each segment, measurement is conceivably possible at two levels: (1) measurement of the total man-machine system outputs for comparison to expected mission goals, and (2) measurement of human operator activity in relation to system outputs [Vreuls, 1974].

TABLE I: NAVIGATION MANEUVERS AND SEGMENTS

SCENARIO	MANEUVERS	SEGMENTS
Navigation	Point-to-Point Flight	Dead Reckoning Contact (visual) Inertial Radar Terrain Mapping
	Nap-of-the-Earth	Very Low Map Interpretation
	Airways-Radio	VOR TACAN ADF
	Off Airways-Radio	Area Navigation
	Over Water	LORAN Celestial Global Positioning System

Source: Vreuls and Cotton [1980]

II. STATEMENT OF PROBLEM

The need for aircrew performance measurement and assessment has long been recognized across all aviation communities. Performance measurement produces information needed for a specific purpose, such as the evaluation of student performance or the identification of aircrews needing training. Unfortunately, the assessment of aircrew proficiency in those skills associated with advanced flying training still depends largely on subjective evaluations by qualified instructor pilots (IPs) and instructor B/Ns (IB/Ns), supplemented with analytically-derived somewhat objective mission performance metrics, e.g., bombing scores [Obermayer, et al., 1974]. An economically acceptable means of objectively measuring behavioral skills in the operational or crew training environment has continued to be a critical problem in the FRS, due mainly to a "nice to have" and nonessential outlook towards any performance measurement scheme other than the traditional "always done this way" method of subjective ratings. A Department of Defense review of tactical jet operational training in 1968 commented: "The key issue underlying effective pilot training is the capability for scoring and assessing performance . . . in essence, the effectiveness of training is dependent upon how well performance is measured and interpreted [Office of Secretary of Defense, 1968]."

Despite the key issue of scoring and assessing performance being identified, current aircrew performance measurement by IPs and IB/Ns during simulated missions in the A-6E WST is all subjective in nature, although the 2F114 simulator has a current objective performance measurement capability. More details on current aircrew performance measurement by instructors in the A-6E WST will be given in Chapter VI.

A. PROBLEM STATEMENT

Current student performance measurement and assessment in the A-6E WST by an instructor is entirely subjective in nature. The A-6E WST has the capability to objectively measure student performance, but is not being utilized in this fashion. Measuring performance is the key to training effectiveness. Objective measurement for aviation training programs has been prescribed by higher authority. Effective performance measurement by using objective methods is vital to establishing performance criteria, the effective utilization of the simulator, instructor effectiveness, aircrew skill identification and definition, and the Instructional Systems Development (ISD) systems approach to training. The problem that must be addressed is designing a performance measurement and evaluation system for the B/N during radar navigation that will incorporate the characteristics of objective performance measurement and still retain the judgement and experience of the instructor as a valuable measuring tool. This thesis will focus upon performance measurement as a system with definable components that interact and produce information necessary for the successful identification of the skill level of the student in regards to navigating the A-6E aircraft. As a result, the A-6E WST can be utilized more effectively, instructors can become more effective in teaching students critical combat skills, and students can complete FRS training being identified at a minimum skill level and "mission ready" for full-system radar navigation in the A-6E aircraft.

B. THE IMPORTANCE OF OBJECTIVE PERFORMANCE MEASUREMENT

A number of factors have contributed to the emerging role of objective aircrew performance measurement in both actual flight and simulators of military aviation units. Generally, this role has developed through an increased awareness of the advantages associated with objective measurement, and the several basic disadvantages of the subjective evaluation method, as outlined in Chapter I.

The remainder of this section will outline both potential and actual necessities for objective performance measurement of aircrew in the training environment. Beginning with a study of Department of Defense policy toward aircrew performance and evaluation methods, the benefits of objective performance measurement are discussed in regards to: standards establishment, increased simulator, instructor, and training effectiveness, aircrew skill level identification and definition, and lastly, ISD requirements.

1. Policy Guidance

Several studies offer guidance with respect to the issue of using more objective measurement techniques in air-crew training. This guidance supports the development and utilization of objective performance measurement as an adjunct or complement to current subjective ratings. In 1968, the Department of Defense review previously cited found that "subjective evaluation was the technique in general use in training programs observed" and had been since before World War II [Office of Secretary of Defense, 1968]. The study went on to comment, "Judgement and experience can be helped by quantitative analytical methods" and that the application of such methods serves three purposes:

(1) They make it necessary to identify the standards of performance desired for each of the many events the pilot must learn.

(2) They determine how many practices or trials a student must accomplish, on the average, to meet the desired standard.

(3) They tell the manager how much improvement he normally may anticipate with each additional practice or trial.

This study concluded: "The services should apply objective evaluation techniques where currently feasible in parts of existing training programs . . ." and "where valid performance data in aircrew training programs can be recorded and stored, quantitative analytical methods should be used to assist the commander in making decisions concerning revising and adjusting the course."

A study by the Comptroller General of the United States (General Accounting Office) in 1973 to the Congress on the use of flight simulators in military pilot training programs stated, "Simulators could also be used to more accurately measure pilot proficiency by using systematic grading procedures." A lack of standardized grading instructions which did not show performance tolerances for the Navy was noted. Conclusions reached were:

> Objective grading of pilot proficiency using simulators would provide more consistent and accurate results for many phases of flight training and eliminate the possibility of human bias and error associated with the current evaluation method . . . simulator grading accurately evaluates pilot proficiency for certain flight maneuvers.

2. <u>Establishment of Standards</u>

The performance criteria, or standard, is a statement or measure of performance level that the individual or group must achieve for success in a system function or task [Office of Secretary of Defense, 1968]. When performance standards are established on the basis of subjective experience and expertise, the result in most cases will be inadequate. When standards are set too low, some risk is incurred with degraded system effectiveness. When set too high, costly overtraining is the result [Riis, 1966; Office of Secretary of Defense, 1968; Campbell, et al., 1976; Deberg, 1977; Rankin and McDaniel, 1980]. The establishment of a standard or baseline of performance is an important result of objective performance measurement. The Department of Defense review in 1968 stated, "Reliable measures of pilot performance against validated standards is the keystone for determining how much instruction and practice is required to attain desired levels of skills [Office of Secretary of Defense, 1968]." Even though the importance of performance standards is recognized, some concern by the A-6 FRS instructors has occurred about establishing operational standards for aircrew performance, due to possible misuse, incorrect adaptation in the training program, or insufficient assessment before implementation [Campbell, et al., 1976]. This issue will be addressed in Chapter VII.

3. Effective Use of Simulators

Objective performance measurement increases the effective use of simulators. When performance measures and criteria are defined, inputted, and monitored by an automatic system requiring little instructor intervention, other training and teaching functions of the simulator may be used by the instructor; thus an increase in the effective use of simulators occurs [Danneskiold, 1955; Knoop, 1968].

4. Effectiveness of Instructors

The major impact of an effective measurement method on the instructor during a simulator mission would be to free him from monitoring dials, Cathode Ray Tubes (CRTs), and lights for aircrew performance measurement and evaluation. Due to the complexity of the A-6E WST, the simulator instructor is humanly unable to monitor and interpret in real-time all pertinent performance information during a training mission.

Objective measurement techniques would relieve the instructor of these monitoring duties, enabling more time for teaching and evaluating those aspects of human performance that are inaccessible by objective methods [Smode and Meyer, 1966; Knoop, 1968; Vreuls, et al., 1974; Kemmerling, 1975; Carter, 1977; Charles, 1978, Semple, et al., 1979]. When performance standards are established by objective methods, instructor judgements can be made more reliable and valid by confining the instructor's judgement to evaluating performance without the additional burden of establishing and adjusting personal standards [Office of Secretary of Defense, 1968]. Efficiency of instructor utilization may be achieved by allowing instructors more flexibility in identifying and assisting students who are found to be deficient from objective performance measurement feedback of criterion levels reached [Carter, 1977; Deberg, 1977; Kelly, et al., 1979]. Such objective information might also provide instructors diagnostic information about their own performance as a teacher after seeing patterns of strengths and weaknesses in their students [Kelly, et al., 1979].

5. Effectiveness of Training

Many factors influence simulator training effectiveness, including: simulator design, the training program, students, instructors, and the attitude of personnel towards the simulator [Tindle, 1979]. Smode and Meyer [1966], in a review of Air Force pilot training, concluded: "The development of objective scores to be used in simulator training would represent a major step toward improving the effectiveness of pilot training programs." Other notable results of objective measurement can also contribute to increased training effectiveness. The real-time feedback of performance measurement to the student is essential to the fundamental concept of knowledge of results, a prerequisite to any learning process. Quantitative feedback, in turn, allows the student and instructor to determine the student's individual strengths and weaknesses in performing the mission, which may then be concentrated on by the instructor for remedial training [Smode and Meyer, 1966; Obermayer, et al., 1974, Deberg, 1977; Carter, 1977; Pierce, et al., 1979; Kelly, et al., 1979]. Modifications in training methods, course content, and sequence of course material could be more accurately assessed by the FRS training officer [Vreuls and Obermayer, 1971; Pierce, et al., 1979; Kelly, et al., 1979]. Student progress within a training program can be more accurately monitored, culminating with the introduction to the fleet of an "operationally capable" or "mission ready" aircrew member at minimum cost [Riis, 1966; Campbell, et al., 1976; Pierce, et al., 1979].

6. Skill Identification and Definition

The employment of objective measures in simulator training will enable the identification and definition of critical combat skills of mission ready aircrews. The precise definitions of "current" and "proficient" and the quantitative measurement of these concepts continues to be a major problem in both training and fleet environments today [McMinn, 1981]. Objective performance measurement requires the definition of "proficient" as a prerequisite to quantification of performance [Pierce, et al., 1979].

7. Instructional Systems Development

Instructional Systems Development is currently being applied to military flight training systems. The approach requires extensive analysis of the specific training to be accomplished, the behavioral objective for each task to be trained, and the level of proficiency required [Vreuls and Obermayer, 1971]. In support of ISD, measures and a measurement system are necessary to: (1) perform analyses of systems in their operational environments, (2) establish quantitative instructional standards, (3) provide an index of achievement for each behavioral objective, and (4) evaluate alternative instructional content, approaches, and training devices [Vreuls and Obermayer, 1971; Obermayer, et al., 1974; Deberg, 1977; Prophet, 1978; Kelly, et al., 1979].

When a state-of-the-art flight simulator is available to an ISD team, it should be the basic medium around which the course is organized [Prophet, 1978]. Campbell, et al. [1976] applied the ISD process to the design of an A-6E aircrew training program and used the A-6E WST for a large part of student training. The study concluded that "difficulty was experienced in applying the ISD process to the development of Specific Behavioral Objectives (SBOs) and criteria test statements, due to the lack of documented quantitative standards of performance." In a review of U.S. Navy fleet aviation training program development, Prophet [1978] reviewed the A-6E ISD application as well as three other major ISD efforts for various aircraft. The results of that study concluded that one ". . . major shortcoming was in the area of performance measurement and evaluation," and recommended measurement as a possible future area for improvement to the ISD model.

The need for incorporating objective performance measurement methods has been addressed. The methodology for the introduction of objective performance measurement into the A-6E WST for B/N performance will now be discussed.

III. METHODOLOGY

A. PROCEDURE

The methodology used in formulating a model to measure B/N performance during radar navigation in the A-6E WST was based on an extensive literature review and an analytical task analysis of the B/Ns' duties. Figure 1 illustrates the approach taken in this report. After selection of the mission, scenario, and segment of interest, the review concentrated on aircrew performance measurement research, which emphasized navigation, training, and skill acquisition. A model was then formulated to show the relationship among student skill acquisition, performance evaluation, and the radar navigation task. This hybrid model, discussed in Chapter V, was improvised by the author specifically to illustrate difficult concepts of aircrew performance measurement and evaluation. The literature review identified different approaches taken in using performance measurement from a systems point of view, some of which were integrated and applied to the current situation.

An in-depth task analysis of the B/N was performed with the purpose of generating candidate performance measures for operator behavior. Skills and knowledge required to perform the radar navigation maneuver were identified and a mission time line analysis was conducted to identify tasks critical to performance. A model was formulated of the A-6E crew-system

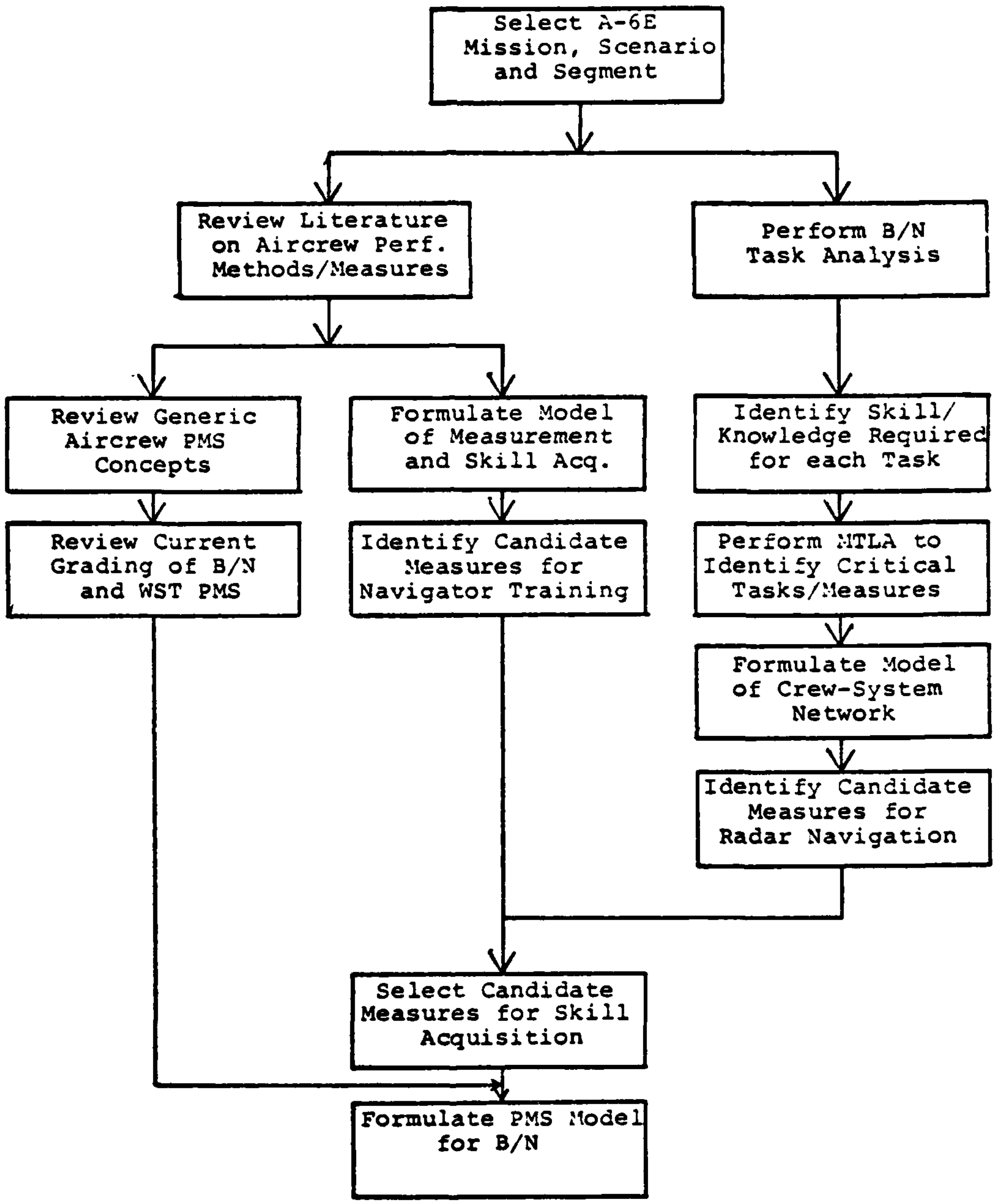


Figure 1. Methodology Flow Diagram.

interactions to illustrate the complexity involved in measuring B/N performance. After defining the purpose of measuring B/N performance, candidate performance measures were identified for possible use in measuring B/N performance.

Candidate performance measures from the literature and the task analysis were compared and measures were selected that met the criteria of face validity, ease of use, instructor and student acceptance, and appropriateness to the training environment. These candidate measures were then compared to current B/N student performance measurement and generic performance measurement systems. The result was a performance measurement system for the B/N during radar navigation in the A-6E WST. Evaluation models were then investigated; a sequential sampling decision model was selected for B/N performance evaluation.

B. ASSUMPTIONS

This section will present some underlying assumptions that are necessary for performance model development and implementation, beginning with a discussion on the necessity of the A-6E WST to realistically duplicate the A-6E CAINS aircraft in both engineering and mission aspects. The unique role of the pilot during the radar navigation mission in the A-6E WST is discussed with respect to his contribution to measuring the B/N's performance. A discussion of the literature review in respect to the relationship between results from pilot studies and navigator performance is presented, followed by a discussion of the need for the existence of a mathematical relationship between measurement and operator behavior. Finally, an assumption is stated concerning the relationship between motivated and experienced aircrew and high skill levels. These discussions follow.

1. Simulator Fidelity

It is assumed that the A-6E WST represents to a satisfactory degree those elements of the A-6E CAINS aircraft such that the A-6E WST aircrew is confronted with a realistic "duplication" of the operational situation and that the aircrew should be required to perform as they would in actual aircraft flight. Given this assumption, training and performance evaluation can be effectively achieved in the simulator for most B/N activities.

2. Pilot and B/N Relationship

The A-6E effectiveness in terms of crew-system output is a function of pilot and B/N crew coordination. Because of the major role of the B/N's activities in achieving the desired mission success during A-6E radar navigation, it is assumed that any variability within the A-6E system that can be measured and attributed to the pilot will be small. In effect, the pilot's function within this mission, scenario and segment will be to "fly system steering," which, for the most part, is the result of the B/N's performance as a navigator and systems operator.

3. Literature Review Results

Most of the literature in the area of aircrew performance measurement has for the most part concentrated on the pilot for performance measurement and evaluation. For similar missions, scenarios and aircrew tasks, it is assumed that what was a significant result in terms of performance measurement for a pilot will be much the same result as that for a navigator. This assumption does not include the psychomotor domain of human operator performance entirely, but does draw some parallels from pilot research results to the navigator. Although each position within the aircraft is somewhat different, many similarities are assumed to exist in terms of operator output, man-machine system output, and measures of effectiveness.

4. Mathematical Description of Behavior

It is assumed that a mathematical relationship exists between some aspects of operator behavior and performance measurement and evaluation. Most likely, for the multidimensional aspects of behavior, a multi-descriptive mathematical result would best describe that behavior in valid and reliable terms. Objective performance measurement relies for the most part on numerical and statistical analysis of operator and system outputs. Thus, this assumption is necessary for the utilization of objective performance measurement to measure and evaluate the B/N's control movements.

5. Experience and Skilled Behavior

When properly motivated and presented with a realistic simulated flight mission with the representative flight tasks, highly experienced ("fleet qualified") aircrews are assumed to exhibit skilled behavior of an advanced stage or high level that is characterized by minimum effort and consistent responses ordinarily found in actual aircraft flight for the same mission. The demonstrated performance of highly skilled aircrew, under this assumption, allows for the establishment of performance standards from which comparisons can be made to populations of aircrew that are less than highly skilled. Both the problem of motivated behavior and establishment of performance standards will be discussed in Chapter VII.

IV. THE FUNDAMENTAL NATURE OF PERFORMANCE EVALUATION

This section is not intended to be a definitive exposition on performance measurement and evaluation theory. However, certain basic concepts of performance measurement and evaluation need to be defined and explained so that a common understanding of subsequent chapters will occur with minimum confusion. This material is approached with a logical time-dependency sequence, beginning with measurement theory, and ending with some desirable characteristics of a total performance measurement and assessment system in the training environment.

Four major areas of performance evaluation will be discussed in this section: measurement considerations, criteria considerations, performance measurement considerations, and performance evaluation considerations. The main purpose is to show that measurement and criteria are needed before the evaluation process begins. Measurement considerations include the definition and purpose of measurement, types of measures, levels of measurement, transformations, measurement accuracy, reliability and validity of measurement, and the selection of initial measures for man-machine performance. The area of criteria considerations addresses the definition and purpose of criteria, types of criteria, characteristics of criteria, establishing criteria, sources of criterion error, measures of effectiveness, and selection of criteria. Performance measurement considerations include other aspects of performance measurement such as: subjective versus objective measures, combining measures, overall versus diagnostic measures, individual versus crew performance, and training measures. The last area of this section, performance evaluation considerations, shows how evaluation depends upon measurement and criteria, and discusses the definition and purpose of performance evaluation, types of evaluation, accuracy of evaluation, evaluating individual and group differences, and the characteristics of evaluation.

The reader already familiar with the above material may wish to skip ahead to the next section. Others not familiar will need the theory to aid understanding of subsequent chapters.

A. MEASUREMENT CONSIDERATIONS

1. Definition and Purpose of Measurement

Measurement is information about performance for a specific purpose, such as whether or not a student is "mission ready" to navigate a particular aircraft [Vreuls and Cotton, 1980]. Unfortunately, this definition leaves open the serious question of quantification; just what and how do you measure and then transform the raw data into useful information? In elemental measurement theory, measurement involves the assignment of a class of numerals to a class of objects, where the class of objects becomes human behavior, the class of numerals must be defined, and some sort of rules for assigning the numerals to the objects must exist [Lorge, 1951; Forrest, 1970]. Measurement then becomes an abstract concept of "mapping" a class or set of numerals to a class or set of human behaviors or performance, but this concept then becomes more quantifiable in nature. All measurements are estimates of the true value or actual amount of the human behavior possessed at a given point in time [Smode, et al., 1962]. The difficulty in the measurement of human behavior increases when the important aspects of the behavior being measured are more qualitative than quantitative in nature [Glaser and Klaus, 1966]. Measurement requires the action of observation, where behavior is noticed or perceived and recorded. Glaser and Klaus [1966] and Lorge [1951] noted that some observations can be made directly, involving perceptions of the behavior or of the behavior's properties, where other observations can only be estimated from inferences about the behavior, or its properties from its effects on other system components. The steps to measurement, as outlined by Forrest [1970], include:

(1) Determine the specific object or aspect to be measured.

(2) Locate or expose the particular object or aspect to view.

(3) Apply a measurement scale.

(4) Express the results as a dimension.

Measurement must precede the activity of performance evaluation, which is the process of interpreting the results of measurement and comparing them to an established standard. Measurement in the training context serves a variety of functions which emphasize either achievement (present knowledge or skill level) or prediction (expected performance under specified conditions). Several specific purposes of training performance measurement as outlined by Smode, et al. [1962], Buckhout and Cotterman [1963], Glaser and Klaus [1966], Vreuls and Obermayer [1971], Farrell [1974], and Vreuls, et al. [1975] are enumerated below:

(1) Prediction of future performance of a student for a specified future operational setting.

(2) Present performance evaluation of knowledge, skill level, or performance level of a student.

(3) Learning rate evaluation at several points in a training program to provide a basis for judging a student's present stage of learning for subsequent advancement to the next training phase.

(4) Diagnostic identification of strengths and weaknesses of a student so that additional training may occur.

(5) Training effectiveness resulting from the nature and extent of training syllabus or course material changes.

(6) Criterion information necessary for defining what constitutes successful or proficient performance.

(7) Functional requirements for future training equipment.

(8) Selection and placement of the student with an achieved level of proficiency to a position or mission with a required minimum proficiency level.

2. Types of Measures

Measurement is a process of producing raw data in the form of measures (parameters or variables) as a result of observing human behavior in a man-machine system. Measures are quantities which can take on any of the numbers of some set and which usually vary along a defined continuum, or scale [Knoop, 1968]. The classification of measures is commonly done by using the characteristics of the measures themselves. Using taxonomies developed by Smode, et al. [1962], Angell, et al. [1964], and Vreuls and Obermayer [1971], measures may be grouped into several major classes with a collection of like measures within each class. The major classes are listed and briefly defined below:

(1) Time periods in output or performance.

(2) Accuracy or correctness of output or performance.

(3) Frequency of occurrence or the rate of repetition of behavior.

(4) Amount achieved or accomplished in output or performance.

(5) Quantity used or resources expended in performance in terms of standard references.

(6) Behavior categorization by observers (subjective measurement).

(7) Condition or state of the individual in relation to the task which describes the behaviors and results of that behavior on system output.

This classification produced approximately 83 measures within the seven major classes and are listed completely in Smode, et al. [1962]. A more recent taxonomy by Mixon and Moroney

[1981] grouped objective only aircrew performance measures by the following six major classes:

(1) Physiological outputs from the operator.

(2) Aircraft systems, instruments or equipment.

(3) Man-machine system output within the operating environment.

(4) Time periods in output or performance.

(5) Frequency of occurrence or the rate of repetition of behavior.

(6) Combined overall measures.

These measures were obtained from an extensive literature review of aircrew performance measurement spanning the years 1962-1980. Table II lists over 180 measures within each major class. It is interesting to note that all of these measures were obtained from actual aircrew performance field or laboratory results, in contrast to previous listings of proposed or candidate measures. Some measures listed in Table II will be used as candidate measures for B/N performance during radar navigation in the WST (see Table XI).

a. Distributions of Measures

The process of measuring continuous and discrete human behavior results in a sample of measures that are estimators of the actual operator behaviors in the system being examined. The usefulness of the measures becomes apparent when they are used to describe the behavior as "good" or "bad" when compared to a reference or standard measure (criterion). Statistical techniques are used to transform raw measures into

TABLE II: A TAXONOMY OF MEASURES

PHYSIOLOGICAL OUTPUTS	
	Biochemical analysis (blood)
	Cardiovascular
	Communications/Speech
	Electroencephalogram (EEG)
	Electromyogram (EMG)
	Eye movements
	Finger tremor
	Galvanic Skin Response (GSR)
	Metabolic rate
	Perspiration weight loss
	Pupillography
	Respiration
	Temperature
	Time of sleep
	Urinary catecholomines
	Visual Evoked Potential (VEP)
AIRCRAFT SYSTEMS	
	ADI displacement
	Aileron
	Aircraft gross weight
	Angle of attack
	Approach glideslope display error
	Approach localizer display error
	Automatic Direction Finder (ADF)
	Autopilot vertical tracking error
	Ball angle
	CDI error
	Collective
	Control stick
	Cyclic
	DME
	Elevator
	Engine Pressure Ratio (EPR)
	Engine RPM
	Flaps
	Flight director error
	Fuel flow
	Heading
	Inclinometer

Source: Mixon and Moroney [1981].

TABLE II (Continued)

AIRCRAFT SYSTEMS (Cont'd)	Landing gear
	Pedal (helicopter)
	Radar altimeter error
	Rotor RPM
	Rudder
	Speed brake
	Tail rotor position
	Throttle
	Thrust attenuator
	Thumbwheel inputs
	Trim
MAN-MACHINE SYSTEM OUTPUT	
	Acceleration
	ACM plane of action (X,Y,Z)
	Aircraft/boom oscillations
	Airspeed
	Altitude
	Altitude (pitchout)
	Altitude (radar)
	Approach angle error
	Approach centerline error
	Approach glideslope error
	Approach range
	Checklist errors
	Circular bomb error
	Closing speed
	Course error
	Crosstrack error
	Deviations from ideal flight path
	Dip to target error (ASW)
	Distance traveled
	Dive angle at bomb release
	Drift
	Emergency detections
	Energy Management Index (EMI)
	Ground speed
	Ground track
	Landing aim point
	Landing attitude
	Ldg. dist. to ideal touchdown point

TABLE II (Continued)

<u>MAN-MACHINE SYSTEM OUTPUT</u> (Cont'd)	Ldg. dist. to runway threshold
	Ldg. height at runway threshold
	Landing result (flare, bolter, ...)
	Lateral acceleration
	Lateral velocity
	Mach number
	Maneuvering rate
	Miss distance (ASW)
	Navigational accuracy (LAT/LONG)
	Pitch
	Pitch/roll coordination
	Pointing angle advantage
	Position estimation
	Positional error (formations)
	Power
	Prob. of finding turn point
	Prob. of target acquisition
	Prob. of target detection
	Procedural errors
	Range at target detection
	Range at target identification
	Range at target recognition
	Rate of information processing
	Ratio: carrier accidents
	Ratio: carrier bolters
	Ratio: carrier bolter rate
	Reaction to an event
	Roll
	Sideslip
	Takeoff position error
	Torque
	Tracking error
	Turn errors
	Turn rate
	Vertical acceleration
	Vertical velocity
	Yaw
<u>TIME</u>	Combined total seconds of error
	Defensive time

TABLE II (Continued)

TIME (Cont'd)	Lead time
	Offensive time
	Offensive time with advantage
	Opponent out of view time
	Ratio: offensive/defensive time
	Reaction time to an event
	Time
	Time estimation
	Time of event
	Time of task execution
	Time on criterion
	Time on target
	Time to acquire target
	Time to criterion
	Time to detect target
	Time to envelope
	Time to first kill
	Time to identify target
	Time to recover from unusual attack
	Time to turn
	Time within criterion
	Time within envelope
	Time within flight path
	Time within gun range
	Time within missile range
FREQUENCY	
	No. of aircraft ground impacts
	No. of collisions (formations)
	No. of control losses
	No. of control reversals
	No. of correct decisions
	No. of correct responses
	No. of correct target acquisitions
	No. of correct target classifications
	No. of correct target detections
	No. of correct target identifications
	No. of correct trials
	No. of course corrections
	No. of crossovers
	No. of errors per trial

TABLE II (Continued)

FREQUENCY (Cont'd)	No. of errors to criterion
	No. of false target detections
	No. of false target identifications
	No. of gun hits/kills
	No. of incorrect control inputs
	No. of incorrect decisions
	No. of lost target contacts (ASW)
	No. of missile hits/kills
	No. of overshoots
	No. of refueling disconnects
	No. of qualifying (criterion) bombs
	No. of scorable bombs
	No. of successful unusual attack rec.
	No. of taps (secondary task)
	No. of target detections (no fires)
	No. of target hits
	No. of target kills
	No. of target misses
	No. of times inside criterion
	No. of times off target
	No. of times outside criterion
	No. of turn points found
	No. of turns to assigned heading
COMBINED OVERALL MEASURES	
	Good Stick Index (GSI)
	Landing Performance Score (LPS)
	Objective Mission Success Score (OMSS)
	Trials to criterion

useful forms of information for the comparison or evaluation process. Since all measures are, in effect, random variables in a statistical sense, it is important to determine the family, or distribution population, that characterizes the particular measure being examined before performing any statistical operations. Two such example distributions would be the Exponential (time measures) and the Normal or Gaussian (accuracy measures). Each distribution has preferred statistical summary estimators that use the generated measures (data) in order to best estimate the actual or true operator performance dimension at hand.

b. Error Measures

Accuracy, or error measures, are of special interest in aircrew performance measurement due to the obvious relationship between operator error and aircraft accidents. These unique measures are usually expressed as a measurable deviation of a variable from an established or arbitrary reference point, and have been of great interest to the aviation accident investigation community [Hitchcock, 1966; Ricketson, 1974]. Chapanis [1951] dichotomized errors as basically constant or variable; constant errors indicated the difference between a statistical estimator of a quantity and the true, or expected value, and variable errors are described by a statistical estimator of measure spread or dispersion. That study concluded that variable errors indicated the actual instability of the man-machine relationship and thus were more of a contributing factor to accidents.

c. Aircrew-Aircraft Measurement

A descriptive structure for flight crew performance measurement relating system performance and human behavior to segments of maneuvers which constitute a flight mission was recently developed by Vreuls and Cotton [1980], derived from earlier work by Benenatti, et al. [1962], Knoop [1968], and Vreuls, et al. [1973]. The structure states that a measure must have meaning only when specified fully by the following five determinants:

(1) Measure segment.

(2) System state variable(s) and their scaling.

(3) Sampling rates for continuous variables.

(4) Desired values.

(5) Transformations.

Measure segments are any portions of flight for which the desired behavior of the system follows a lawful relationship from an unambiguously defined beginning to end. Segments are related closely to specific behavioral objectives (from the ISD training approach), are possibly but not required to be the same as a maneuver segment, and defined explicitly by start/stop logic. System state variables, as previously discussed (measures), are scaled so that their entire dynamic range is represented without information loss. Scaling will be discussed along with desired values and transformations later in this section. Sampling rates are the temporal frequency at which a measure is recorded or observed by the measurement system. One sampling rate guideline is to record data faster than the natural frequency response for the specific axis in which the measurement is being made, although others are proposed [Vreuls and Cotton, 1980].

3. Levels of Measurement

In examining the nature of performance measurement, four levels can be distinguished. The level of scale determines the amount of information resulting from the measurement and the mathematical and statistical operations that are permissible within that level. Table III is adapted from Lorge [1951], Siegel [1956], and Smode, et al. [1962], and lists each level and the applicable characteristics associated with each. A brief description of each level as analyzed by Smode, et al. [1962] is discussed below:

a. Nominal measurement scales have units being measured placed into classes or categories. Units placed in the same class are considered equivalent along some dimension or in some respect.

b. Ordinal measurement scales have units assigned a rank order. Nominal categories are now ranked with respect to each other in terms of an "amount possessed" of the quantity being measured, with judgements assessing the amount possessed by the units involved. Rankings can be composed of unidimensional or multidimensional variables. In the latter case, a composite judgement or ordering is performed which essentially places a multidimensional variation on a unidimensional linear scale.

TABLE III: LEVELS OF MEASUREMENT

SCALE	BASIC EMPIRICAL OPERATIONS	MATHEMATICAL GROUP-STRUCTURE	PERMISSABLE STATISTICS
NOMINAL	Determination of equality Provides identity only	Permutation group $x'=f(x)$	Frequencies Mode Contingency coefficient
ORDINAL	Determination of greater or less Provides both identity and order	Isotonic group $x'=f(x)$ (where $f(x)$ means any monotonic increasing function)	Frequencies; Medians; Modes; Percentiles; Contingency coefficient; Rank order coefficient Nonparametric statistical tests
INTERVAL	Determination of equality of intervals or differences Provides identity, order, and addivity	General linear group $x'=ax+b$	Frequencies Means Modes Standard deviation Percentiles Rank order coefficient Nonparametric statistical tests Parametric statistical tests
RATIO	Determination of equality of ratios Provides identity, order, and addivity with reference to absolute zero	Similarity group $x'=ax$	

Source: Lorge [1951], Siegel [1956], and Smode, et al. [1962].

c. Interval measurement scales have units being measured in equidistant terms. In addition to an indication of not only rank order or direction, there is also an indication of the amount or size of difference between units on the scale. Since the zero point is usually arbitrary, it does not represent complete absence of the property being measured.

d. Ratio measurement scale is an extension of an interval scale with a natural, absolute zero point, and represents the highest measurement level. It is with this level that the most powerful statistical tests of significance are available when evaluating performance measures.

The determination of a parameter or measure of performance should include the units of scaling, i.e., 0 to 640 knots (air speed), and 0 to 64000 feet (pressure altitude). Without a clear definition of the scaling units, the improper use of statistical operations or tests may occur, causing the measurement of performance to provide irrelevant or erroneous results.

4. Measure Transformations

Measures are observed, recorded, and usually subjected to transformation, which is any mathematical, logical, or statistical treatment designed to convert raw measures into usable information [Vreuls and Cotton, 1980]. When measures are discrete, the transformation may be the actual value, absolute value, or a tolerance band. When measures are continuous, transformations may include means, variances, frequency content, departures from norms, or several others as listed in Table IV.

The relationship between the distribution of a measure or estimate and transformations is well known to statisticians but sometimes not very clear to others. For a given population distribution, there exists unbiased and consistent estimators (transformations) that will best describe the true value or quantity that is being estimated. Indeed, some transformations are not applicable for a given population, and in a sense are useless. The interested reader is referred to Mood, et al. [1974] for a detailed analysis of applicable transformations for a known population distribution.

5. <u>Accuracy of Measurement</u>

Measurement produces <u>measures</u> which are sample estimators of the true value or actual amount of the quantity possessed. Accuracy of measurement refers to how close an estimator is to the true value. All measurement systems are subject to accuracy problems as discussed below.

a. Measuring aircrew behavior is confounded by the systematic influence of the total operating system, since the measures obtained are frequently determined to some extent by the performance of other components in the system [Glaser and Klaus, 1966].

b. Any statistic, or known function of observable random variables that is itself a random variable, whose values are used to estimate a true quantity, is an estimator of the

TABLE IV: COMMON MEASURE TRANSFORMATIONS

TIME HISTORY MEASURES

- Time on target
- Time out of tolerance
- Percent time in tolerance
- Maximum value out of tolerance
- Response time, rise time, overshoot
- Frequency domain approximations
    - Count of tolerance band crossing
    - Zero or average value crossings
    - Derivative sign reversals
    - Damping ratio

AMPLITUDE-DISTRIBUTION MEASURES

- Mean, median, mode
- Standard deviation, variance, range
- Minimum/maximum value
- Root-mean-squared error
- Absolute average error

FREQUENCY DOMAIN MEASURES

- Autocorrelation function
- Power spectral density function
    - Bandwidth
    - Peak power
    - Low/high frequency power
- Bode plots, fourier coefficients
    - Amplitude ratio
    - Phase shift
- Transfer function model parameters
    - Quasi-linear describing function
    - Cross-over model

BINARY, YES/NO MEASURES

- Switch activation sequences
- Segmentation sequences
- Procedural/decisional sequences

Source: Vreuls and Cotton [1980].

the true quantity and may be subject to biased or inconsistent properties. For all distributions of measures that have been identified during measurement, there exists at least one unbiased and consistent estimator that will be closer to the true value of the quantity being measured than any other estimator. Using the incorrect estimator for a known population will, in effect, result in avoidable measurement inaccuracies [Krendel and Bloom, 1963; Mood, et al., 1974].

There is no simple way to assure measurement accuracy, but several techniques to improve the accuracy of measurement may be incorporated and follow.

c. Scope of Measurements

Accuracy will increase as a result of increasing the inclusiveness or completeness of the measures to include all relevant behavior and system components. Skilled performance in an aircraft normally involves complex behaviors that require a wide scope of measurement to accurately estimate the existing true performance level [Smode, et al., 1962].

d. Number of Measurements

For most situations, the greater the number of observations involved in deriving an estimator, the closer the estimator wil be to the true value [Mood, et al., 1974]. Increasing the number of observations when there is a large variability in the individual measurements will reduce the variability and minimize the effects of random or chance variations which may occur from measurement to measurement

[Smode, et al., 1962]. Large sample sizes also are desirable when establishing standards or references (criteria) when evaluating performance [Krendel and Bloom, 1963].

e. Controlled Conditions of Measurement

By insuring the desired measurement conditions are controlled, accurate measurement may be improved. This may be done by defining those factors which are to be present and varied systematically, maintaining uniformity of conditions during measurement in order to reduce bias and unwanted variability, and insuring the intended measurements are taken correctly [Smode, et al., 1962].

6. <u>Reliability of Measurement</u>

Reliability refers to the accuracy of measurement or the consistency or stability of the recorded and statistical data upon repetition [Glaser and Klaus, 1966; Knoop and Welde, 1973; Grodsky, 1967; Thorndike, 1951]. When the dispersion, or spread, of measures obtained from one individual on a particular task is large, the measures lack reliability, and any statistics that are formed from the measures that are used in evaluation will be incapable of differentiating consistently among individuals who are at different skill levels. If the measures are precise, resulting in a statistic that is stable, an individual would receive exactly the same evaluation score each time his performance was measured [Glaser and Klaus, 1966; Thorndike, 1951].

a. Computation of Reliability

Thorndike [1951] cautioned: "There is no single, universal, and absolute reliability coefficient for a test. Determination of reliability is as much a logical as a statistical problem." Several methods of approximating reliability have since been proposed or utilized in aircrew measurement. Smode, et al. [1962] classified reliability expressions as either absolute or relative and suggested using the standard error of measurement (absolute measure of precision), coefficient of internal consistency (uses single set of observations), coefficient of stability (measure agreement over time), and the coefficient of equivalence (agreement between two like measures) as statistical computations of reliability. Glaser and Klaus [1966] dichotomized reliability assessment into two methods: test-retest and alternate form, and recommended computing reliability by using the statistical correlation coefficient. Some recent statistical techniques - Winsorization and trimming - may provide a better reliability approximation than was previously possible. Winsorization and trimming involve removing the effects of a large variability in a measure sample, with virtually no loss in information [Dixon and Massey, 1969]. These techniques would appear to be quite useful for reliability calculations, although their actual use in aircrew performance measurement has not yet been demonstrated.

b. Sources of Unreliability

The sources of measurement accuracy problems, as previously discussed, are also sources of unreliability. Other sources inherent in the measurement of human behavior, which hinder reliable measurement of aircrew performance, are from Glaser and Klaus [1966] and Thorndike [1951], and include the following:

(1) The environment in which performance is being measured influences measurement variability. Differences in weather, amount of turbulence, wind direction and velocity, unexpected noise, equipment malfunction, and extreme temperatures are factors contributing to unreliability.

(2) Equipment used for measurement or personnel participating in performance assessment are sources of unreliability.

(3) The complexity of the behavior being evaluated influences unreliability. Since the behavior being measured and evaluated involves many dimensions of performance, and any individual's skill level may fluctuate considerably from one dimension to the next, each component dimension is susceptible to all previously discussed sources of unreliability that have been described above.

(4) The performance of the individual being assessed fluctuates as a function of temporary variations in the state of the organism. Some factors frequently involved that decrease measurement reliability are: individual

motivation, emotional state, fatigue, stress, test-taking ability, and circadian rhythm.

c. Improving Reliability of Measurement

In any training situation, some degree of reliability in the measure of the ability being trained is necessary if any evidence of improvement is required. By reducing chance factors or variability in measurement, reliability can be improved. The techniques for improving the accuracy of measurement, as previously discussed, also contribute towards improving reliability. Knoop and Welde [1973] suggested other factors to improve reliability that are listed below:

(1) Calibration of performance measurement equipment should be conducted on a continuing basis.

(2) Software processes involved in data collection, reduction, conversion, analysis, and plotting should be validated and monitored to avoid data loss.

(3) Accurate records of flight conditions and mission requirements should be maintained to facilitate measurement interpretation.

7. Validity of Measurement

Validity is the degree to which the measurement or evaluation process correctly measures the variable or property intended to be measured [Knoop and Welde, 1973]. In regards to the evaluation process, validity has two aspects: (1) relevance or closeness of agreement between what the test measures and the function that it is used to measure, and (2) reliability or the accuracy and consistency with which the test measures whatever it does measure in the group with which it is used [Cureton, 1951]. In the training environment, a performance test is a stimulus situation constructed to evoke the particular kinds of operator behavior to be measured or assessed. The validity of a performance test is established by demonstrating that the test results reflect differences in skill levels of the performance being assessed [Glaser and Klaus, 1966].

a. Types of Validity

Four types of validity which have applicability to performance measurement in general have been described by Smode, et al. [1962], McCoy [1963], and Chiles [1977] as follows:

(1) Predictive validity refers to the correlational agreement between obtained measures and future status on some task or dimension external to the measurement and requires statistical operations for evaluation.

(2) Concurrent validity refers to the correlational agreement between obtained measures and the present status of the units being measured on some task or dimension external to the measurement and also requires statistical computation.

(3) Content validity is based on expert opinion and is evaluated by qualified people determining if the measures to be taken truly sample the types of performance or subject matter about which conclusions will be drawn.

(4) Construct validity is a logical process where the emphasis is on trait, quality, or ability presumed to underlie the measures being taken. While the measures themselves do not reflect directly the performance to be evaluated, they are accepted to be valid on the basis of logical considerations and related empirical evidence.

b. Validity of Measurement in the Simulator

Without empirical or judgmental evidence, the use of full-scale state-of-the-art flight simulation provides maximum face validity, where the performance evaluation situation in the simulator appears to duplicate the actual task of flying or navigating an aircraft [Alluisi, 1967; Chiles, 1977]. Bowen, et al. [1966], in an experiment using twenty A-4 pilots to study and assess pilot proficiency in an Operational Flight Trainer (OFT; device 2F62), found that:

> For measures taken in the OFT to be valid, the task set to the pilot should be multiple in nature and have a considerable difficulty level equivalent to the more difficult levels found in actual flight. In this manner, the pilot is more likely to display his skills in the same pattern of priority (i.e., time-sharing, attention-shifting, standard-setting, etc.) as he does in actual flight.

This conclusion is also supported by Kelley and Wargo [1968] and in terms of the relevance component of validity as previously discussed by Cureton [1951].

c. Improving Validity of Measurement

A high degree of validity is essential to the effectiveness of any measurement system. Improving validity can be achieved by increasing either or both relevance and reliability. Relevance may be increased by reproducing the simulation situation to closely resemble that of the actual aircraft itself, or by simulating the task or mission being performed more closely to the actual task or mission environment. Reliability improvements were discussed in the previous section.

d. Relationship of Validity and Reliability

Validity has been described as having aspects of relevance and reliability. Reliability is the consistency or self-correlation of a measurement while validity is its correlation with some independent standard or reference from that which is measured [Kelley and Wargo, 1968]. A given performance measurement can be highly reliable yet not have validity [Smode, et al., 1962; Kelley and Wargo, 1968]. However, an unreliable test cannot have practical validity, i.e., a measurement that does not even correlate well with itself will not correlate well with other measurements [Kelley and Wargo, 1968; Steyn, 1969]. In performance measurement, high validity must be combined with high reliability; this combination means that a highly skilled operator consistently must achieve a higher performance evaluation result than a less skilled operator [Smode, et al., 1962; Kelly, et al., 1979]. If the performance evaluation occurs during the actual task instead of a simulated task, the question of validity reduces simply to the question of reliability, as perfect relevance will have been achieved [Cureton, 1951].

Because of the unique relationship of validity and reliability, it is generally easier and more efficient to improve the reliability of a measure than to raise its validity. On the other hand, if the validity of a measure appears promising, improving reliability is preferred to using a reliable measure which has lower validity [Glaser and Klaus, 1966].

8. Selection of Initial Measures

After the identification and selection of a desired mission, scenario, flight segment, and human operator with the behavior of interest, performance associated with these requirements can be specified and defined. The initial selection of appropriate measures has been a major problem, as evidenced by the lack of concordance in recent aircrew performance measurement research [Mixon and Moroney, 1981]. Unless aircrew performance measurement empirical results are collected and standardized, an analytical approach must be taken when initially selecting which measures best describe the performance being examined. Some optimum balance exists between the "measure everything that moves" philosophy and the measurement of a few measures with apparent face validity. The initial selection of any measures, however, should be guided by both the purpose of the measurement and the man-machine system as well as the facility for collecting and processing the data. The following criteria for initial measure selection are provided by Meister and Rabideau [1965], Parsons [1972], Greening [1975], and Vreuls and Wooldridge [1977]:

(1) Relevance - the measures should be pertinent to the purpose of measurement.

(2) Quantifiable - measures should be in the form of numerals on a ratio scale for statistical analysis, except where only subjective collection is feasible.

(3) Accessibility - measures should not only be observable but easily collectable, preferably by electronic means.

(4) Operational utility - a measure that has relevance and accessibility in both the aircraft and simulator environments.

(5) Efficiency - a measure with utility at minimum cost of collection and transformation into usable information.

(6) Content validity - a positive correspondence between the performance measure and what is known about the underlying behavior.

(7) Reliability - collection of more measures or samples of a measure than planned would offset the likelihood of electronic data collection equipment failures.

(8) Dependence - the availability of human or automatic data observers and collectors limit what measures are feasible to collect.

(9) Objectivity - where possible, automatic data observation and recording is preferred to human observers and recorders.

(10) Usable - measures collected should be usable for either evaluation information or supportive to evaluation results.

(11) Acceptable - measures that are used by instructors in the operational environment must be consistent with their expert judgement.

(12) Collection criteria - data must be accurate and precise to what is known about the underlying behavior.

Once the performance measures of interest are identified and selected, they should be defined, as a minimum, in terms of the descriptive structure as previously outlined by Vreuls and Cotton [1980]. Some consolation in not selecting the appropriate measures for performance is offered by Knoop [1968], "Determining exact criteria [standards] and performance measures for virtually any flight task or mission is an accomplishment yet to be made."

B. CRITERIA CONSIDERATIONS

1. Definition and Purpose of Criteria

Criteria are standards, rules, or tests by which measures of system behavior are evaluated in terms of success or failure, or to some degree of success or failure. The purpose of human performance criteria is to provide standards or baselines for evaluating the success or failure, goodness or badness, or usefulness of human behavior [Knoop, 1968; Vreuls and Cotton, 1980; Cureton, 1951; Steyn, 1969; Davis and Behan, 1966; Shipley, 1976; Buckhout and Cotterman, 1963]. The criterion is a measuring device which is not generally or readily available, but a device which should be constructed from the beginning for each particular situation [Steyn, 1969; Christensen and Mills, 1967]. Criteria should not only define the unique manner in which the operator should perform a task, but should define the performance objectives of the entire man-machine system [Demaree and Matheny, 1965; Connelly, et al., 1974].

2. Types of Criteria

The classification of criteria can be accomplished from a measurement standpoint; begin with the smallest known entity and end with the "ultimate" quantity that may exist. Several types identified are listed below:

(1) Parametric referent or standard of performance which is sought to be met by the operator or system. Example: maintain 500 feet of altitude [Demaree and Matheny, 1965; Shipley, 1976].

(2) Parametric limit about the parametric standard within which the operator or system is required, or seeks, to remain. Example: maintain plus or minus 100 feet while at 500 feet altitude [Demaree and Matheny, 1965; Shipley, 1976].

(3) System component criteria which distinguishes the relationship between system components and system output. Example: "least effort" measured from the pilot in relation to maintaining altitude [Krendel and Bloom, 1963; Uhlaner and Drucker, 1964].

(4) Test criterion used to evaluate overall human ability, usually expressed as a single overall measure. Example: subjective judgement of instructor for a student as to "pass" or "fail" [Marks, 1961].

(5) Ultimate criteria are multidimensional in nature and represent the complete desired end result of a system. This type of criterion is impossible to quantify due to the multidimensional nature of the system's purpose, and hence, is a theoretical entity that must be approximated. Example: Any aircraft's mission [Cureton, 1951; Smode, et al., 1962; Uhlaner and Drucker, 1964; Steyn, 1969; Shannon, 1972].

It may be noted that all five types of criteria can be quantified or approximated in some manner, with decreasing accuracy as the ultimate criteria level is reached. Obtaining direct measures of the ultimate criteria for a complex system is seldom feasible. This is particularly true in military systems where such criteria would be expressed in terms of combat effectiveness or effectiveness in preventing a potential aggressor from starting a conflict [Smode, et al., 1962]. Therefore, it becomes apparent that we must select intermediate criteria (types one through four above) in evaluating skilled operator behavior.

3. Characteristics of Good Criteria

Using actual criteria as approximations of the ultimate criteria can be accomplished by several methods that will be discussed in a later section. Although there is no certain method that will lead to the specification of good criteria, there are some considerations that can be taken into account which are discussed below:

(1) A good criterion is both reliable and relevant [Smode, et al., 1962; Krendel and Bloom, 1963; Cureton, 1951; Grodsky, 1967; Steyn, 1969].

(2) Criteria must be comprehensive in that the utility of the individual being evaluated is unambiguously reflected [Steyn, 1969].

(3) Criteria should possess selectivity and have ready applicability [Krendel and Bloom, 1963].

4. Other Criteria Characteristics

Steyn [1969], in a review of criterion studies, noted that performance measures under simulated conditions can at best serve as substitute criteria. This observation reflects the engineering and mathematical model of reality represented by the simulator that can, at best, approximate an aircraft and its systems. Shipley and Gerlach [1974] measured pilot performance in a flight simulator (T4-G) and found that differences in pilot performance outcomes varied as a function of the difference in criterion limits that were established, with the relationship between criterion limits and tracking performance found to be a nonlinear one. The multidimensionality of a criterion was also noted by Steyn [1969], Cureton

[1951], and Connelly, et al. [1974]. These latter two studies observed that multiple criteria must exist for a single task since different operator action patterns having no single feature in common could conceivably obtain the same desired system output.

5. Establishment of Criteria

Criteria may either be derived from regulatory requirements, system operating limits, knowledge of common practice, or empirical studies [Vreuls and Cotton, 1980]. When criteria are established analytically, some caution must be taken. Campbell, et al. [1976], in designing the A-6E TRAM training program using ISD methodology, observed that:

> A standard or criterion of performance for that terminal behavior must also be established . . . at a level comparable to the earlier established operational standards. These latter criteria, however, while reflecting an acceptable level of behavior, imply a repeatability, that . . . whenever they are performed, that some acceptable level will be attained.

Criteria derived from objective, empirical techniques are preferable to analytical methods [Steyn, 1969]. Regression analysis, discriminant function analysis, multivariable regression, and norm or group referencing are but a few of the empirical approaches to establishing criteria [Connelly, et al., 1974; Danneskiold, 1955; Dawes, 1979]. No matter which method is used, criteria must be defined and are necessary for the evaluation process.

6. Sources of Criterion Error

As previously mentioned, a good criterion is one that is both reliable and relevant. Reliability, as previously defined, implies that what constitutes successful performance will be resistant to the effects of chance factors. Relevancy refers to the validity of the actual or approximated criterion to the ultimate criterion. By definition, the ultimate criterion is completely relevant. Sources of criterion error can then be identified in terms of reliability and relevance. Smode, et al. [1962], lists some significant sources of criterion error below:

(1) Low reliability, as previously mentioned.

(2) Irrelevancy or the lack of relation of the actual criterion with respect to the ultimate or "ideal" criterion.

(3) Contamination of the criterion by the presence of factors or ingredients in the actual criterion which do not in fact comprise the ultimate criterion.

(4) Distortion in the criterion caused by errors arising from assigning incorrect weights to the separate factors that comprise the actual criterion (combining criteria is discussed below).

7. Measures of Effectiveness

Aircraft missions are all multidimensional in nature. This means that every mission can be divided into usually one overall goal or purpose (i.e., destroy the target, deliver the supplies, rescue the survivors, etc.), with several subgoals (safety, minimize susceptibility, timeliness, etc.). Since missions are multidimensional, the operator effort in the form of mental and physical action (performance) becomes multidimensional. The multidimensional nature of skilled aircrew performance, in turn, requires that several criteria, all of which are relevant for a particular activity, be defined and used [Smode, et al., 1962]. Each of these criteria must be operationally defined, theoretically quantifiable, and collectively give a reasonable portrayal of operator and system performance. Typically, one may wish to bring these component criteria together in an overall effectiveness measure - a single "measure of effectiveness" for the system being investigated. The process of combining criteria into a single composite measure of effectiveness is one of the most difficult tasks to undertake in any field of research, and has been the focus of continuous investigation in the science of Operations Research for decades [Hitch, 1953; Morris, 1963; Steyn, 1969; Lindsay, 1979; Dawes, 1979].

A Measure of Effectiveness (MOE) is a quantifiable measure used to compare the effectiveness of the alternatives in achieving the objective, and must measure to what degree the actual objective or mission is achieved [Operations Committee, Naval Science Department, 1968]. For the particular situation of measurement and evaluation of an aircrew member's performance in an aircraft, criteria can be thought of as "alternatives," each of which has an individual effectiveness for each performance component of an aircrew member, who is accomplishing an overall objective - the mission. MOE's have been applied in economics, management, and military problems - just about any area where a decision based on information from system performance has to be made.

Combining criteria into an MOE can be accomplished either analytically or statistically. Most methods concentrate on assigning weights that are either determined on the basis of "expert" opinion or statistical treatment [Smode, et al., 1962; Steyn, 1969]. In a review consisting of numerous criterion weighting studies, Steyn [1969] concluded that "it would appear that the most acceptable approach [to weighting criterion variables] would be to identify the job dimensions clearly and unambiguously and to use these pure dimensions as criteria to be predicted independently."

There is no established procedure for combining criteria into a single overall MOE. Lindsay [1979] and Smode, et al. [1962] offer some suggestions to approach the problem:

(1) Look at the big picture. Examine what is to be done with the results of the aggregation. Determine how the numbers will be used, and in what decisions. (Usually one finds that this has not been thought out in advance.) It may be that all that is really needed is the identification of satisfactory systems.

(2) If possible, aggregate subjectively. Give the subcriteria values to the decision-makers or their advisers and let them subjectively determine how effective the systems are.

(3) Recognize that one is defining, not approximating. The development of a formal scoring system should be done with the awareness that a definition of system effectiveness is being made. The procedure developed should include reference points, diminishing marginal returns, and avoid substitutability except where appropriate.

(4) Sub-criteria should be weighted in accordance with their relevance to the ultimate criterion.

(5) Sub-criteria which repeat or overlap factors in other sub-criteria should receive a low weight.

(6) Other things being equal, the more reliable sub-criteria should be given greater weight.

The unique situation of an aircrew flying an aircraft for a specific mission and the necessary determination of sub-criteria for evaluating the overall accomplishment of that mission requires further research of an analytical and empirical nature. The relationship among altitude, airspeed, operator activity, and the hundreds of other system variables that comprise the total system must be compared to mission success in quantifiable terms. Since "mission success" is multidimensional and may not be totally measurable and quantifiable, some analytical approaches toward combining criteria into an overall MOE appear to be feasible, notwithstanding the possible use of empirical methods to describe some aspects of the process.

8. <u>Selection of Criteria</u>

Criteria have been defined, their purpose established, some types identified, and some characteristics discussed. Since a large number of criteria may exist to evaluate a particular system, some selection in the way of "trade-offs" may be necessary [Davis and Behan, 1976]. More than one criterion may describe the same dimension of performance whereas another carefully selected criterion may accurately describe more than one performance dimension. Reducing the number of criteria that are relevant, reliable, and practical into a feasible and usable set that can accurately and consistently evaluate the performance of an aircrew and the accomplishment of their mission is extremely difficult at present and will probably remain as an unsolved future problem in the Human Factors field unless specific research is undertaken to attack it. In the meantime, some general guidelines for selecting criteria as discussed by Hitch [1953] and Smode, et al. [1962] are listed below:

(1) Selection of any criteria should always be consistent with the highest level or type of criterion associated with the system mission.

(2) Specify the activity in which it is desired to determine successful and skillful performance.

(3) Consider the activity in terms of the purpose or goals, the types of behaviors and skills that seem to be involved, the relative importance of the various skills involved, and the standards of performance which are expected.

(4) Identify the elements that contribute to succesful performance and weight these elements in terms of their relative importance.

(5) Develop a combined measure of successful performance composed of sub-criteria that measure each element of success and are weighted in accordance with the relative importance of each.

The definition, computation, combining and selection of criteria is perhaps the most difficult problem encountered by researchers investigating complex man-machine systems [Osborn, 1973; Krendel and Bloom, 1963]. The importance of criteria is once more emphasized, as stated by McCoy [1963], "You must define the criterion precisely and accurately before interpreting any measures used in investigating a system." Christensen and Mills [1967], in quoting earlier work done by H.R. Leuba, stated:

> There are many ludicrous errors in quantification as it is practiced today, but none quite as foolish as trying to quantify without a criterion. It is awkward enough to quantify the wrong thing when a criterion exists, but it is a sham of the most unprofessional sort to quantify in the absence of a criterion.

C. PERFORMANCE MEASUREMENT CONSIDERATIONS

1. Subjective Versus Objective Measures

As previously discussed in Chapter I, subjective and objective measures are not dichotomous, but rather represent a continuum of performance measurement. At one extreme of the continuum, a human observer mentally records actual performance during a specified mission, and uses his perceptions to form a judgement or degree of success rating as to how skillful the operator was in achieving the system objectives. This extreme is the subjective method of measurement and evaluation. At the other extreme of the performance measurement continuum, automatic digital computers sense, record, transform, analyze, and compare actual man and system performance to statistically established criteria and form a complete set of performance data or information to be used by the decision-maker (instructor or training officer) in evaluating the skills and abilities of an operator. This other extreme is the objective method of measurement and evaluation.

Each method of performance measurement has advantages and disadvantages, as were discussed previously, and will not be repeated here. Objective measures and measurement have become more feasible and less costly to implement for aircrew performance than in previous decades, and the method has established itself as a very powerful and useful model for describing actual human behavior [Mixon and Moroney, 1981].

2. Combining Measures

What has been discussed previously with respect to combining criteria also applies to combining performance measures into a single overall index of skill level or proficiency. As Smode, et al. [1962] indicated:

(1) Measures should be weighted in accordance with their relevance to the criterion.

(2) Measures which repeat or overlap factors included in another measure should receive a low weight.

(3) Other things being equal, the more reliable measures should be given greater weight.

In combining performance measures, it is often possible to determine quantitatively the interrelationships among the performance measures and the relationship between each measure and the actual or immediate criterion [Smode, et al., 1962]. A single overall measure or score composed of numerous performance measures along different dimensions of system behavior is highly desirable in any performance measurement and evaluation system, due to the use of the total score in determining overall performance when compared to a criterion. A single score or estimate of total performance, when compared to a criterion or MOE, provides the necessary information for evaluation that determines goodness or badness, success or failure, and usefulness of human performance [Buckhout and Cotterman, 1963].

Combining performance measures, like criteria, can be performed by either analytical or empirical methods which commonly assign weights to each measure which are then mathematically combined into a single proficiency score. Analytical methods employ the judgement of experts for situations usually involving complex man-machine systems where definitive and quantifiable measures of output are not available [Glaser and Klaus, 1966; Marks, 1961]. Empirical methods of combining aircraft system performance measures with relative weightings into a single score were reviewed by Vreuls and Obermayer [1971]. Among some of the methods from that study for developing multidimensional algorithms were: factor analysis, multiple discriminate analysis, linear-weighted algorithm, nonlinear (threshold) model, energy maneuverability model, time demand, recursion models, and empirical curve fit. The interested reader is referred to that study for more detail on each model and the circumstances in which it was employed.

In summary, separate performance measures of different behavioral dimensions are combined by various analytical and empirical methods into a single overall or composite score with the idea that when the score is high, as compared to a predetermined criterion or MOE, it indicates "good" or "successful" performance, and when low, indicates "poor" or "unsuccessful" performance [Cureton, 1951]. Difficulties in how to combine the measures into a single overall score leads to preservation of the behavior dimensions and a "vector" of measures.

3. Overall Versus Diagnostic Measures

Overall measures of skilled performance, as previously discussed, along with total system output measures (e.g., bomb-drop accuracy, number of targets hit, fuel consumed) are beneficial in assessing total system performance but are seriously lacking in diagnostic information of potential value to the trainee [Buckhout and Cotterman, 1963; Kelley and Wargo, 1968; Bergman and Siegel, 1972]. Overall scores tell nothing about the operator's performance on various specific tasks which are involved in flying an aircraft on a mission, but are highly useful for performance evaluation.

Diagnostic measures are the same measures that result from performance measurement before any combining operations take place. These measures identify certain aspects or elements of a task or performance in specific skill areas and provide useful information on strengths and weaknesses in individual skills [Smode, et al., 1962]. Since they are concerned with smaller and more precisely defined units of behavior, they are easier to measure by objective methods.

It thus appears that overall and diagnostic measures are contradictory but both essential. For the training environment, where a student is learning skills necessary to perform a task, both measures are valuable for what information they provide, as discussed above. Using the two together in a complementary fashion was perhaps best stated by Smode, et al. [1962], "A prime value of an overall measure is the support it provides in evaluation since diagnostic measures alone are difficult to interpret without some terminal output measure of performance." Kelley and Wargo [1968] recommended that performance be measured in each dimension, evaluated separately by comparison to specific and predefined criteria, and then combined into an overall total score, so the trainee can receive feedback relating to his relative performance on the various dimensions of his task, as well as on his overall performance. More recently, Vreuls and Wooldridge [1977] described multivariate statistical modeling techniques that are powerful enough to provide measures for diagnosis, and yet also provide single measures that could be combined into an overall score.

The qualities of overall and diagnostic measures have been described and their relationship has been discussed. Within the training environment, using one without the other appears to cause a decrease in the quality of information available to the individuals who most need it: both student and instructor. Therefore, for the design of a system to measure student B/N performance during a radar navigation mission, it appears advantageous to employ the use of both overall and diagnostic measures in a mutually beneficial manner that will provide the maximum amount of accurate information for the purposes of training situations.

4. Individual Versus Crew Performance

One of the assumptions of this thesis is that the variability of the total contribution of the pilot in the conduct and successful accomplishment of the radar navigation mission is small enough to essentially be ignored when measuring the performance of the B/N during the same mission. This assumption was based on the major role played by the B/N during radar navigation, the unique design of the A-6E CAINS navigation system, and the radar navigation mission itself. Although it is recognized that any successful accomplishment of a mission depends to some degree on the crew interactions and coordination, the actual measurement of the interaction and coordination was beyond the current scope of study, and will be left for future investigation and research.

The unique problem of measuring crew coordinated performance has been the focus of much research, but the question of what "crew coordination" is remains unanswered [Mixon and Moroney, 1981]. Smode, et al. [1962] provided a detailed discussion on the problems and approaches taken in measuring aircrew coordination, and concluded that measured interaction and communication were good for differentiating "good" and

AD-A098 776 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 5/9

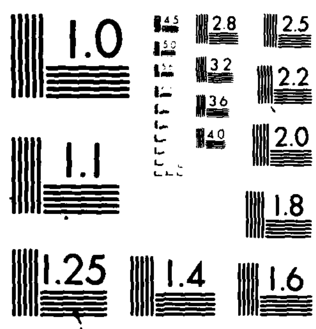
A MODEL TO MEASURE BOMBARDIER/NAVIGATOR PERFORMANCE DURING RADA--ETC(U)

MAR 81 T R MIXON

UNCLASSIFIED NL

2 of 4

AD A098776



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

"bad" crews. Good crews reduced individual interaction and communication to a minimum so that more time was available to devote effort to performing the individual tasks associated with accomplishing the mission. This conclusion does not acknowledge the inherent difficulties involved in objectively measuring individual interaction and communications. Until further research uncovers objective, valid, reliable, and practical methods of measuring crew coordination, this area is perhaps a measurement function best delegated to a human observer (instructor).

5. Measures and Training

The importance of overall and diagnostic measures in the training environment has been previously discussed. Measures for the evaluation of performance are related in some degree to the stages of training. Early in training, when skilled behavior is made up largely of familiarization with the task and basic knowledge of procedures, measurement may consist of more familiarization and procedure-related measures. Late in training, when skilled performance has become more or less automatic, measurement becomes more difficult due to the highly cognitive and covert nature of the skilled behavior [Fitts, 1965; Glaser and Klaus, 1966]. In this case, measurement becomes more indirect than direct.

Designing any measurement system within the training environment requires a detailed understanding of the training process and its relationship to performance measures, in addition to an explicit understanding of the basic nature of the skills involved in performing the task. The latter subject will be discussed in Chapter V.

D. PERFORMANCE EVALUATION CONSIDERATIONS

Although the purpose of this thesis is to design a system to improve current performance measurement techniques for the FRS B/N student, some mention must be made of performance evaluation since performance measurement exists as information necessary to evaluate individual and system performance. Without evaluation, little reason exists for the measurement, recording, and storage of performance measures. This section will briefly outline current evaluation methods and the characteristics of evaluation itself.

1. Definition and Purpose

Being consistent with the previous discussions on performance measurement and criteria, performance evaluation is simply the process of identifying and defining performance criteria and then comparing the criteria to performance measures produced by performance measurement. All performance evaluation requires some comparison between a standard and an estimate of what the standard truly represents [Angell, et al., 1964; Demaree and Matheny, 1965]. The purpose of performance evaluation in the training environment is usually multidimensional in nature but all evaluation occurs for the purpose of accurate decision-making by the instructor regarding student performance and by the training officer for effective training control. On the instructor level of evaluation, faulty decision-making due to any performance evaluation involves two possible errors: Type I and Type II, as found in Table V.

TABLE V: PERFORMANCE EVALUATION DECISION MODEL FOR THE INSTRUCTOR

		REALITY: UNSKILLED	REALITY: SKILLED
DECISION:	Student has not acquired skill to perform task	Correct decision	Type I (α)
	Student has acquired skill to perform task	Type II (β)	Correct decision

The tangible effects of a Type I error are possible increased costs due to overtraining, an inefficient training flow of students, and a demotivated student. On the other hand, a Type II error may result in increased costs due to an aircraft accident and the loss of human life. This example illustrates the important role that performance measurement and subsequent evaluation plays in providing accurate information necessary for correct decision-making by the instructor.

2. <u>Types of Performance Evaluation</u>

Based on the purpose of the evaluation, evaluation may be divided into two general types: aptitude and achievement. According to Marks [1961], if the purpose is to predict the capacity of a trainee to absorb training and perform a task, the evaluation is called an aptitude test. If the purpose is to tell how well the trainee has absorbed training or can perform the task, the evaluation is called an achievement measure. When considering achievement measures, it is possible to distinguish three basic kinds:

(1) Proficiency tests require the individual to answer questions about his job or about some content knowledge area related to his job.

(2) Performance tests involve controlled observation of an individual actually performing his job.

(3) Rating methods use the opinion of someone who has actually seen the man's performance on the job.

For details on the characteristics, advantages, and disadvantages of each kind of achievement measure, the interested reader is referred to Marks [1961].

A model is anything that represents reality. Two performance evaluation models that are utilized in achievement measures are norm-referenced testing and criterion-referenced testing.

a. Norm-Referenced Testing

Norm-referenced testing involves the use of norm-referenced measures in evaluating performance. Norm-referenced measures compare the performance of an individual with the performance of other individuals having similar backgrounds and experience [Glaser and Klaus, 1966; Knoop and Welde, 1973; Danneskiold, 1955]. The stability of a norm-referenced measure is highly dependent upon sample size. Too small a sample can yield measures of central tendency and variability that poorly approximate actual population values [Glaser and Klaus, 1966; Danneskiold, 1955].

b. Criterion-Referenced Testing

Criterion-referenced testing uses criterion-referenced measures for making an evaluation of performance. These measures involve a comparison between system capabilities and individual performance [Glaser and Klaus, 1966]. Such measures indicate whether an individual has reached a given performance standard [Knoop and Welde, 1973]. The standard for criterion-referenced measures may be determined either by analysis, subjective judgements by a panel of experts, or numerous successful performances as sampled from a large population [Knoop and Welde, 1973].

c. Criterion- Versus Norm-Referenced Testing

A recent article by Swezey [1978] reviewed and described the relative advantages and disadvantages of criterion-referenced and norm-referenced testing, from which the conclusions are cited below:

> Content validated criterion-referenced tests, which are derived from appropriate job, task, or training analyses, often provide the best available measure of performance; particularly in objectives-oriented situations. It is often the case that no better criterion exists upon which to validate the instrument.

Other researchers have supported this conclusion, especially in the field of aircrew training performance measurement, and it is perhaps a more feasible alternative to the more traditional and less efficient method of norm-referenced testing [Knoop and Welde, 1973; Waag and Eddowes, 1975; McDowell, 1978; Uhlaner and Drucker, 1980].

3. Accuracy of Evaluation

The accuracy of evaluation is dependent upon the accuracy of measurement, the accuracy and relevance of the criteria, and the evaluation conditions. Since the accuracy of measurement and criteria have already been addressed, this section will be limited to evaluation conditions.

During any evaluation, several sources of contamination, or bias, as discussed by Danneskiold [1955] and Glaser and Klaus [1966], may affect the performance evaluation of individuals, and are listed below:

(1) In performance testing, one individual may naturally perform better than another during the examination situation, even though both may actually possess the same skill level.

(2) The sequence and construction of the simulated mission test may cause some individuals to respond in a way that is dependent only on the test sequence and construction.

(3) Judgemental errors occur whenever individuals are used to observe performance, due to prejudices and stereotypes formed by the observer.

(4) Evaluating condenses performance dimensions into a compact and meaningful unit, where in the process some information is lost.

(5) Observed performance is only a sample of the total skills and knowledge of the individual.

The accuracy of evaluation may be increased by improving either measurement accuracy, the accuracy and relevance of criteria, or the evaluation conditions. For evaluation conditions, the common method of eliminating some of the five bias factors mentioned above is to increase objectivity in measurement and to standardize the test conditions [Glaser and Klaus, 1966].

4. <u>Evaluating Individual and Group Differences</u>

The measurement of differences in individual performance is highly desirable in a training situation. As training progresses, the performance of individual trainees gradually approaches the desired minimum skill level required for system operation. The accurate evaluation of what skill level the individual actually possesses during the training process is necessary for efficient training control and for efficient instruction [Glaser and Klaus, 1966]. Several methods have been developed to identify which performance measures can best discriminate among individuals at various ability levels; these will be discussed in Chapter VII due to their applicability in designing the measurement system at hand [Parker, 1967; Buckhout and Cotterman, 1963; Thorndike, 1951].

Measuring group differences, as opposed to individual differences, is more suitable for the purposes of treatment evaluation, such as the training method, length of instruction, and design of displays and controls, and will not be addressed in detail here. Interested readers may consult Glaser and

Klaus [1966] or Moore and Meshier [1979] for further discussions of measuring methods for group differences.

5. Characteristics of Evaluation

Some characteristics and considerations that contribute to improving the evaluation process are listed below:

(1) Repeatability of a measure implies that a specified score achieved today represents the same level of performance as it did at a previous time (temporal invariance) [McDowell, 1978].

(2) Sensitivity of a measure occurs when a measure reliably changes whenever the operator's performance changes [Grodsky, 1967; Kelley and Wargo, 1968; Knoop and Welde, 1973].

(3) Comprehensiveness of the measures employed in covering as wide a range of flying skills as possible [Ericksen, 1952].

(4) Interpretability of measures and evaluation results [Demaree and Matheny, 1965; Waag and Eddowes, 1975; McDowell, 1978].

(5) Immediately available measures and scores to provide the student with knowledge of results [Buckhout and Cotterman, 1963; Demaree and Matheny, 1965; Welford, 1971; Waag and Eddowes, 1975; McDowell, 1978; Kelly, 1979].

(6) Economical considerations require that evaluation be constrained by cost and availability of personnel, yet adequate at a minimum level for the purpose at hand [Marks, 1961; Demaree and Matheny, 1965].

(7) Standardization of test conditions and environments enables performance to more accurately reflect true operator skill [Ericksen, 1952; Marks, 1961; Smode, et al., 1962; Demaree and Matheny, 1965].

This list of desirable characteristics of evaluation is not all-inclusive but does provide some foundation for examining existing evaluation systems for those properties that are in consonance with system and evaluation goals.

V. THE NATURE OF THE BOMBARDIER/NAVIGATOR TASK

A. INTRODUCTION

Navigating an A-6E TRAM aircraft during a low altitude, non-visual air interdiction mission is perhaps one of the most demanding and complex tasks expected of navigators today. The aircraft must avoid rough or mountainous terrain while traveling narrow corridors between geographical turn points, which must be crossed with pinpoint accuracy at predesignated times. Literally hundreds of individual steps or procedures are involved in navigating the aircraft, each of which contribute in some dimension to attaining the mission objective. Figure 2, adapted from Obermayer and Vreuls [1974], shows the crew-system interactions in the A-6E aircraft. The pilot controls the aircraft and manages aircraft flight systems while receiving visual and auditory navigational information from the Vertical Display Indicator (VDI) and B/N, respectively. As it can be noted in Figure 2, the B/N manages the navigational equipment, processes large amounts of concurrent information, and makes critical decisions regarding the navigational accuracy of the aircraft. At any one time, the B/N may be executing tasks that are dichotomous, sequential, continuous, monitorial, computational, or decisional in nature. At all times he is serving as the systems manager.

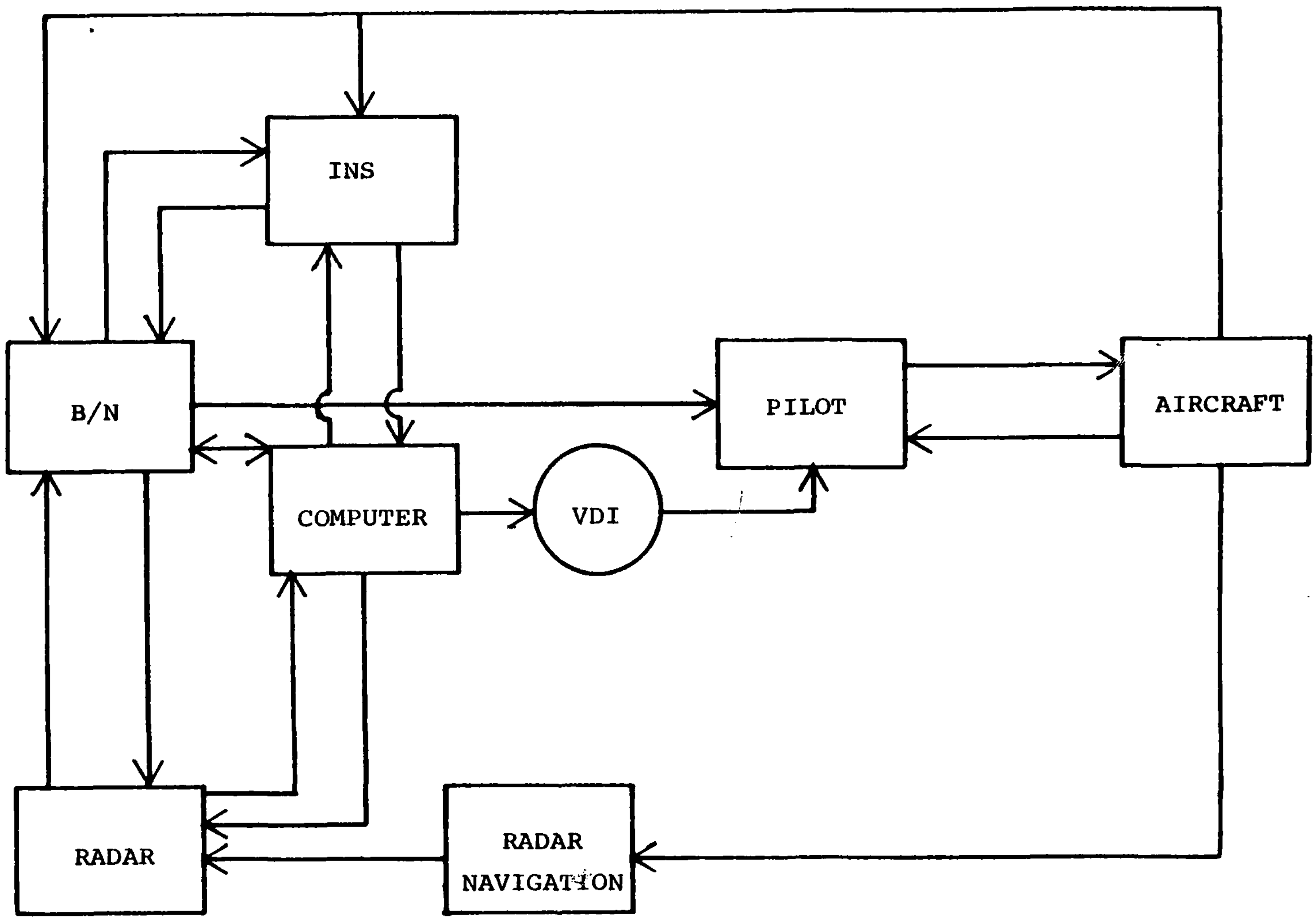


Figure 2. A-6E Crew-System Network.

One of the more difficult subtasks is radar scope interpretation. This activity involves the recognition of the relationship between specific symbols or patterns of symbols on a flight chart with the specific returns or patterns of returns on the radar scope [Beverly, 1952]. The success of this identification subtask depends largely upon the quantity and quality of a priori information about the target that was available to the radar navigator [Williams, et al., 1960]. This subtask may be performed while the B/N is monitoring the Inertial Navigation System (INS), observing computer-generated navigational information displays, and informing the pilot about current equipment status. It is an axiom among student B/Ns that "if you are sitting there perceiving that everything has been done, you are getting behind."

This section will define and describe the nature of the B/N's tasks in terms of: (1) the physical variables of the aircrew-aircraft system, and (2) the complex skills and abilities of a perceptual, psychomotor, and cognitive nature. The importance of operationally describing and systematically classifying the B/N's tasks from a behavioral point of view is that such an analysis may point to areas where measurement of performance is both desirable and feasible, and may indicate the relationships of individual tasks to overall mission success [Smode, et al., 1962; Vreuls, et al., 1974]. The tool used to define and describe the tasks of the B/N will be a task analysis.

B. TASK CONSIDERATIONS

1. Task Definition

A task is one or more activities performed by a single human operator to accomplish a specified objective [Connelly, et al., 1974]. In the aviation training environment, a navigation task is the successful action of the navigator in response to visual, aural, vestibular, and tactile information concerning the actual and desired values of a particular parameter (or more than one parameter) associated with navigating the aircraft, usually after completing a lesson or series of lessons [Demaree and Matheny, 1965; Anderson and Faust, 1974].

2. Classification of Tasks

There are numerous task classification methods, all of which depend on the purpose of describing the tasks and the nature of the tasks themselves. Smode, et al. [1962] classified behavior for performance measurement purposes with the idea of accommodating both diagnostic measures relating to elemental tasks as well as the more global measurements relating to overall system performance. These general behavior classes are listed and defined as follows:

a. Level I - Elemental Tasks

The simplest level of analysis, referring to any homogeneous series of work sequences conducted at one time, or single actions taken toward accomplishing a desired objective. These tasks range from short duration discrete homogeneous acts to longer sequences of routing activity.

b. Level II - Complex Tasks

The composite of activities which involve identifiable sequences of homogeneous activity or recurring single actions and sub-routines in performance. Each complex task is made up of tasks from Level I, involving either the simultaneous and/or sequential integration of combinations of elemental tasks, or the repetition of a single Level I activity over time.

c. Level III - Mission Segments

The segments or phases of performance that are identified in full mission activity. Essentially, a segment is composed of a group of complex tasks (Level II) which are integrated in the performance at this level of description.

d. Level IV - Overall Missions

The major types of missions anticipated for advanced flight vehicles. Each mission is composed of a group of segments of activity which are integrated in the performance at this level of description.

Beginning with overall missions and ending with elemental tasks, these progressive refinements in task specificity allow performance measurement decisions to be made at progressively more detailed levels.

3. Task Relation to Performance and System Purpose

For measurement purposes, it is neither practical nor desirable to measure all possible task conditions which might occur in accomplishing a mission objective [Smode, et al., 1962; Vreuls, et al., 1974]. To be practical, an attempt should be made to simplify the analysis of tasks and remove irrelevant measurement [Vreuls, et al., 1974]. Since measurement is only possible on the basis of specific, observable events, a great deal of investigation and analysis must be accomplished to describe tasks that are representative of accomplishing the mission purpose while at the same time are measurable [Glaser and Klaus, 1966]. Glaser and Klaus [1966] identified two kinds of observable performance that are useful for performance measurement: the behavioral repertory of the operator in the form of verbal and motor actions, and the operator's effects on overall system performance or output. From the measurement of either of these two observed performances, some inference can be made about the operator's level of skill in performing operational and describable tasks.

The intimate relationship between the task of the operator and performance measurement of the operator for the purpose of estimating his level of skill was best stated by Smode, et al. [1962]:

> The behaviors and tasks which are observed and measured necessarily will be a sampling from those which comprise the complete system activity, for it is neither feasible nor necessary to measure everything in order to evaluate proficiency. What one evaluates depends on purpose. Determining those properties of behavior significant to the purpose aids in defining the areas of human behavior for assessment. In the interest of maximizing validity of measurement, this sampling should be guided by the criticality of the tasks and operational segments to mission or system success. As a rule, those tasks should be selected for measurement on which good performance results in mission success and on which poor performance means failure.

As discussed previously, identifying the purpose of the system is important for defining performance standards and for accurate measurement of behavior. Likewise, the purpose of the system helps define those tasks which should be measured, and is essential for discovering what the relationship is between mission tasks performed by the operator and the probability of mission success [Smode, et al., 1962; Buckhout and Cotterman, 1963; Cotterman and Wood, 1967]. Actions that are critical to performance in that they differentiate between success and failure in performance can only be identified properly in terms of the ultimate purpose or goal of the man-machine system [Smode, et al., 1962].

C. RADAR NAVIGATION TASK ANALYSIS

1. Background

A task analysis is a time-oriented description of man-machine interactions brought about by an operator in accomplishing a unit of work with an item of the machine, and shows the sequential and simultaneous manual and intellectual activities of the man operating, maintaining, or controlling equipment, rather than a sequential operation of the equipment [Department of Defense, MIL-H-46855B, 1979]. Miller [1953] presented a more usable definition of a task analysis: the gathering and organization of the psychological aspects of the indication to be observed (stimulus and channel), the action required (response behavior, including decision-making), the skills and knowledge required for task performance, and probable characteristic human errors and equipment malfunctions.

A task analysis is conducted mainly for the design of new systems or for improvements to existing systems and provides basic building blocks for the rest of human engineering analysis [Van Cott and Kinkade, 1972]. The purpose of the task analysis presented here is to improve current performance measurement of the B/N in the A-6E WST, and will be discussed more fully in respect to this purpose in Chapter VII.

There are several methods of conducting a task analysis which are classified as either empirical, analytical, or some combination of both. The empirical methods rely on industrial engineering techniques such as time and motion study [Mundel, 1978] while the analytical techniques involve the use of expert opinions through interviews or questionnaires. Van Cott and Kinkade [1972] advocated seeking information from a wide variety of sources and employing more than one technique in order to adequately describe what an operator actually does in a system.

"A completely developed task analysis will present a detailed description of the component behavioral skills that the accomplishment of the task entails, the relationships among those components, and the function of each component in the total task [Anderson and Faust, 1974]." Since a task analysis involves breaking down a task into behavioral components for the purpose of performance measurement, the question of when to stop subdividing the task is most important. Anderson and Faust [1974] proposed that enough task analysis detail is reached when the intact or component skill is part of the student's entering behavior.

The use of a task analysis for the purpose of performance measurement assumes that behavior can be analyzed in terms of basic components that are conceptually identified in a way that is convenient and agreeable to people and that specific measurement techniques appropriate for the various behavioral components exist [Smode, et al., 1962]. This assumption becomes less theory and more factual in light of research conducted in the helicopter community. Locke, et al. [1965], in a study of over 500 primary helicopter students using the OH-23D helicopter, reported that nearly all complex man-machine maneuvers can be broken down into independent component parts with associated component abilities. A more recent study by Rankin and McDaniel [1980] assessed helicopter flight task proficiency using a Computer Aided Training Evaluation and Scheduling (CATES) system, where flight maneuvers were divided into tasks that were used for performance measurement and evaluation, and were then utilized to determine overall aviator proficiency.

Just as no two task analyses are ever the same, there may be multiple sets of operator behavior possible to accomplish the tasks as described in one task analysis [Fleishman, 1967; Vreuls and Wooldridge, 1977]. This major limitation of using a task analysis to measure performance of an operator reaffirms the idea of measuring all observable system outputs and establishing the relationship among operator actions, system outputs, and mission success or failure. By empirically validating a task analysis in the operational environment and establishing the above mentioned relationships, any limitations imposed by differences in operator strategy on the measurement system may be circumvented.

2. Previous A-6 Task Analyses

a. Naval Flight Officer Function Analysis

In 1972, the Chief of Naval Operations requested the Naval Aerospace Medical Research Laboratory (NAMRL) to conduct a series of investigations analyzing the operational functions of the Naval Flight Officer (NFO) for the purposes of revising NFO training programs and to aid in determining future training equipment requirements and characteristics. Addressing NFOs of P-3B/C, RA-5C, A-6A, EA-6B, E-2C, and F-4B/J aircraft, the investigations determined the roles, duties and tasks performed by the NFO in a given aircraft, the percent of NFOs performing a given task/duty, the time and effort spent on various roles, duties and tasks, and finally, the task criticality.

The study of interest for the purposes of this thesis involves the analysis of the B/N operational functions in the A-6A [Doll, et al., 1972]. The procedure used for the function analysis was based on a method of job analysis developed by the USAF Personnel Research Laboratory at Lackland Air Force Base, Texas. The principal method of analyzing functions was an inventory of activities approach that combined features of the checklist, open-ended questionnaire, and interview methods.

The results of analyzing the A-6A B/N tasks was based on 84 surveys completed by operational B/Ns. Of six major operational roles identified (communication, navigation, tactics, sensors, armament, and system data processing), more time and effort was spent (28 percent) in flight by the B/N performing the navigation role than any other single role. Within the navigation role, five duties were identified: (1) navigate using Inertial Doppler systems, (2) using TACAN, (3) using ADF/UHF-ADF, (4) using visual references/Dead Reckoning, and (5) using radar. Over 98 tasks within those five duties were listed. Amount of time and effort as well as the criticality of each task was recorded and a rank order listing of all tasks for these two categories was presented.

In developing a task analysis for the B/N during radar navigation (presented later in this section), this A-6A function analysis for the B/N shows the importance of navigation in terms of time and effort spent by B/Ns in the operational environment. A measurement system that accurately describes B/N performance during radar navigation would be extremely useful from this standpoint. The time and effort, and criticality rankings of this source were also useful for those tasks that corresponded with the same task or subtask in the current effort, and in developing a performance measurement system that encompassed critical tasks in terms of their contribution to overall mission success.

b. Grumman A-6E TRAM Training Program

Grumman Aerospace Corporation completed a study on the application of ISD methodology to the design of a training program for A-6E TRAM FRS pilots and B/Ns in mid-1976. Comprised of over seven volumes, the study included a task analysis, development of SBOs, media analysis, and formulation of lesson specifications [Campbell, 1975; Campbell and Sohl, 1975; Campbell, et al., 1975; Hanish and Feddern, 1975; Graham, et al., 1975; Campbell, et al., 1977]. The task analysis phase of the ISD process was performed jointly by a team consisting of Navy Subject Matter Experts (SMEs) and Grumman training psychologists, educational specialists, and flight test personnel. Tasks were to be identified based on performance in the operational environment and described in sufficient depth to permit an identification of the underlying skills and knowledge required by the crewmen to perform the task. A hierarchical approach for describing the pilot and B/N behaviors during a mission resulted in three levels of description: major mission events, the tasks which comprise the events, and the steps which describe the incremental actions an aircrewman must take to complete a task.

The first result of the task analysis effort was a comprehensive task listing comprised of over 400 nominal pilot tasks, each with an average of approximately 10 steps; 70 airframe emergency sequences involving an average of 7-10 steps each, 35 system malfunctions, and more than 200 nominal B/N tasks with an average of 10 steps each. The listings represented tasks for which training needed to be conducted at the FRS level. A Task Analysis Record (TAR) form was utilized for each task to ascertain the following: (1) crewman performing task, (2) where training was given, (3) skills and knowledge required by task, (4) conditions under which task is performed, (5) cues involved in performance, (6) aircraft system involved, (7) degree of difficulty, (8) factors in task difficulty, (9) task criticality, (10) factors in performance measurement, and (11) other special factors which impacted on training. Because the TAR was used for the purpose of instructional sequencing and blocking downstream in the ISD process and as an aid in selecting appropriate instructional strategies, it was not published as part of the study.

The actual task analysis appears in the form of an ISD record developed from the TAR and SBOs. Objectives were classified on the basis of eight major taxonomic categories: (1) knowledge, (2) comprehension, (3) discrimination, (4) application, (5) analysis, (6) synthesis, (7) evaluation, and (8) complex performance. This taxonomy was retained for the current thesis task analysis effort and will be defined in Table VII. The ISD record then contained the SBO, task identification data, condition/constraints, performance standard, taxonomic data, a criterion test statement, and test type and format.

The Grumman task analysis effort becomes useful to the current effort of developing a task analysis for the B/N during the radar navigation maneuver, and using that task analysis for the purpose of performance measurement. In this respect, the Grumman study was used as a guiding outline in developing the current task analysis.

Prophet [1978] reviewed past ISD efforts in Navy fleet aviation training program development that included the Grumman A-6E ISD program, and made the following comments in reference to measurement and evaluation for that program:

(1) Methodologies being followed did not necessarily require a systematic treatment of measurement and evaluation.

(2) No discussion of the mechanics of measurement for standards found in SBOs is given.

(3) While a clear recognition of when and where measurement will take place is addressed, no information is given concerning how.

(4) The problems of flight versus non-flight measurement were not discussed.

Although some criticism may be found in the lack of measurement mechanics from the Grumman task analysis effort, it is not a surprising revelation given the purpose of their task analysis. Their effort is still deserving in the light of the complexity of the aircrew and A-6E combination, and this author used their unclassified task analysis material in defining and describing exactly what a B/N does during a radar navigation maneuver.

c. Perceptronics Incorporated Decision Task Analysis

In early 1980, a study designed to identify significant aircrew decisions in Navy attack aircraft was performed by Perceptronics, Inc. for the Naval Weapons Center, China Lake, California. The study selected two mission scenarios that were representative of A-6E and A-7E aircraft: close air support and fixed target attack [Saleh, et al., 1980]. A mission analysis followed by an Aircrew/Avionics Functions Analysis was performed on each scenario. Finally, a decision identification analysis was performed which resulted in a listing of significant decisions in each mission for each aircrew. The study results provided information on decision type, difficulty, and criticality.

Limited use was made of this decision identification analysis due to the scenarios developed and the purpose of the task analysis: decision-making, and some dependence of that task analysis upon the previous efforts by Grumman. Nevertheless, a few decisional tasks were reviewed for use in the current effort.

3. <u>Current Task Analysis for Performance Measurement</u>

Using the research provided by the Naval Flight Officer function analysis, the Grumman A-6E TRAM training program task analysis, and the Perceptronics, Inc. decision task analysis, as discussed previously, a task analysis was performed with the purpose of measuring B/N performance during radar navigation in the A-6E WST. The results of that effort, in the form of a task listing (Appendix A), a task analysis (Appendix C), and a Mission Time Line Analysis (MTLA; Appendix D), are each presented separately below.

a. Task Listing

As shown in Appendix A, the radar navigation maneuver was divided into three segments: (1) after takeoff checks, (2) navigation to the initial point (IP), and (3) navigation to the turn point (TP). The navigation to TP segment (3) was the portion of the A-6E CAINS flight within the scope of this thesis, and was the segment of interest to be later expanded upon in the form of a task analysis and MTLA that will be discussed later in this section.

The following definitions will explain the significance of the symbology within the navigation to TP segment of Appendix C:

(1) Tn - Task number, where the number is represented by "n."

(2) Sn - Subtask number.

(3) (a) - Subtask element, where the element is represented by the lower case letter "a," or other letters.

The nomenclature for actual switches, keys, controls, or buttons in the A-6E CAINS cockpit is underlined throughout the task listing (e.g., <u>Rcvr</u> control). Discrete or continuous settings for each switch, key, control, or button is to the far right of the task, subtask, or subtask element, and is separated by a line of periods.

The choice of language in the form of action verbs for which behaviors are described was a difficult process, due to the lack of standardization in both the science of analyzing tasks and in aircrew performance measurement research. The necessity of employing action verbs that described simple and easily observable activities and were easily identified in terms of performance measurement was paramount to the current effort. A hybrid taxonomy, using 31 action verbs as shown in Table VI, was developed from earlier work by Angell, et al. [1964] that was, in a sense, later validated by Christensen and Mills [1967] in an analysis of locating representative data on human activities in complex operational systems. Using a later study by Oller [1968] in the form of a human factors data thesaurus as applied to task data, the original 50 action verbs used by Angell, et al. [1964] was reduced to a total of 31 action verbs by eliminating redundant synonyms and by using the recommended acceptable action verbs and nouns. Except for the reduction of action verbs (specific behaviors), the remainder of the original taxonomy was preserved. For the convenience of the reader, the 31 action verbs or specific behaviors utilized in the current task analysis are presented as a glossary (Appendix B).

TABLE VI: CLASSIFICATION OF BOMBARDIER/ NAVIGATOR BEHAVIORS

PROCESSES	ACTIVITIES	SPECIFIC BEHAVIORS
Perceptual	Searching for information Receiving information Identifying objects, actions, or events	Checks Monitors Observes Reads
Mediational	Information processing	Initiates Records Uses
	Problem solving and decision-making	Checkouts Compares Continues Delays Determines Evaluates Performs Repeats Selects Troubleshoots
Communication	Communicating	Alerts Informs Instructs
Motor	Simple/discrete	Activates Depresses Places Pushes Throws Sets
	Complex/continuous	Adjusts Inserts Positions Rotates Tunes

Source: Angell, et al. [1964] (adapted)

b. Task Analysis

A task analysis for the specific purpose of measuring B/N performance during radar navigation was performed and is presented as Appendix C. As previously discussed, only segment three, navigation to TP, was examined during the task analysis to limit the scope of this study. Since the concepts of segment and tasks have already been addressed, the seven columns of the A-6E TRAM radar navigation task analysis form in Appendix C will now be explained in detail, using guidance provided by Van Cott and Kinkade [1972], Anderson and Faust [1974], Pickrel and McDonald [1964], Smode, et al. [1962], and Rosenmayer and Asiala [1976]:

(1) Subtask - a component activity of a task. Within a task, collectively all subtasks comprise the task. Subtasks are represented by the letter "S" followed immediately by a numeral. Subtask elements are represented by a small letter in parentheses.

(2) Feedback - the indication of adequacy of response or action. Listed as VISUAL, TACTILE, AUDITORY, or VESTIBULAR and located in the subtask column for convenience only.

(3) Action Stimulus - the event or cue that instigates performance of the subtask. This stimulus may be an out-of-tolerance display indication, a requirement of periodic inspection, a command, a failure, etc.

(4) Time - the estimated time in seconds to perform the subtask or task element calculated from initiation to completion.

(5) Criticality - the relationship between mission success and the below-minimum performance or required excessive performance time of a particular subtask or subtask element. "High" (H) indicates poor subtask performance may lead to mission failure or an accident.

"Medium" (M) indicates the possibility of degraded mission capability. "Low" (L) indicates that poor performance may have little effect on mission success.

(6) Potential Error - errors are classified as either failure to perform the task (OMIT), performing the task inappropriately in time or accuracy (COMMIT), or performing sequential task steps in the incorrect order (SEQUENTIAL).

(7) Skills Required - the taxonomy of training objectives used for the Grumman task analysis was retained and is presented in Table VII [Campbell, et al., 1977]. This concept will be discussed in more detail later in this section.

(8) Performance Measure Metrics - a candidate metric which may best describe the successful performance of the task or a genuine display of the required skills. The types of metrics suggested were classified as TIME (time in seconds from start to finish of task), T-S (time-sharing or proportion of time that particular task is performed in relation to other tasks being performed in the same time period), R-T (reaction time in seconds from the onset of an action stimulus to task initiation), ACC (accuracy of task performance), FREQ (number of task occurrences), DEC (decisions made as a correct or incorrect choice depending on the particular situation and mission requirements), QUAL (quality of a task, especially in regards to radar scope tuning quality), and SUBJ (subjective observation or comprehension of task execution success by an instructor).

Due to the lack of operational data, the task analysis was derived analytically with close attention being paid to consistency with previous A-6 task analysis efforts. The validation of any task analysis can only occur when it is subjected to the operational environment for repeated empirical analysis. Unfortunately, time, cost and system availability constraints precluded the execution of this important phase.

TABLE VII: TAXONOMY OF SKILLS REQUIRED

Knowledge

Technological - Learning "how to" perform a single switch and control configuring procedure. Learning "how to" read meters, digital displays, scopes, lighting displays, etc. In general, learning "how to."

Formal - Learning the meaning of special symbols, acronyms, words, nomenclature, etc.

Descriptive - To describe "what is" and "what was": facts, data, special information about systems, sub-systems, equipment, weapons, tactics, missions, etc.

Concepts and Principles - Fundamental truths, ideas, opinions and thoughts formed from generalizations of particulars.

Comprehension

Understanding the meaning of meter readings, scope, digital and lighting displays. Understanding the switch and control configuring procedure, i.e., the reason for a specified sequence, the reason for a switch or control position, the reason for a verification, etc.

Grasping the meaning of concepts and principles, i.e., understanding the basic principles of infrared and radar detection.

Understanding the meaning of facts, data, specific information, etc.

Discrimination

Distinguishing among different external stimuli and making appropriate responses to them, e.g., scanning gages for out-of-tolerance trends. Also includes the recognition of the essential similarity among a class of objects or events, e.g., classifying aircraft types or radar return images.

Source: Campbell, et al. [1977].

TABLE VII (Continued)

<u>Application</u>

Simple Procedure - A demonstration of a simple learned procedure in the cockpit or simulator requiring not more than simply repeating required switch and control configuring and simple visual verification (i.e., advisory light status).

Complex Procedure - A demonstration of a learned procedure in a cockpit or simulator that requires differentiating or distinguishing between readings on meters, digital displays, and images on video and radar displays and interpreting and applying the meaning of the readings and images.

General - Using learned materials in new and concrete situations (e.g., using rules, methods, concepts, principles, procedures, etc.).

<u>Analysis</u>

A demonstration of a learned process of breaking down material (i.e., data, other information) into its components so that it may be evaluated with respect to crew's safety, mission success, A/C maintenance, etc.

<u>Synthesis</u>

A demonstration of learned process, i.e., putting tactical elements together (e.g., weapons, targets, available systems, A/C capability, etc.) to formulate a mission.

<u>Evaluation</u>

A demonstration of a learned process of assessing or judging a system or situation, based on criteria (i.e., data, rules, available equipment, conditions, etc.) and then reaching a conclusion based on this assessment.

<u>Complex Performance</u>

A demonstration that requires psychomotor skills and/or critical thinking skills usually requiring practice.

As it stands, a reasonable assumption of the existence of some face validity in the current task analysis can be made in the light of the author's operational experience as a B/N in the A-6E CAINS aircraft (over 600 hours) and the dependence of the task analysis upon previous task analysis efforts, even though none of the previous efforts were formally validated by empirical methods. The current task analysis was also informally reviewed by other A-6E B/Ns before finalization of the effort.

The purpose of the task analysis was to improve current performance measurement of the B/N during radar navigation in the WST by providing performance measure metrics (right-hand column of Appendix C) that are possible candidates for describing successful task performance or B/N skill acquisition. Several hundred metrics are available from which a candidate set can be chosen based on the initial measure selection criteria as previously discussed in Chapter IV. From the "performance measure metrics" column of Appendix C, several potential candidate measures were identified and will be combined with potential measures from Table II in Chapter IV and presented as part of the final candidate measure set listed in Table XI (Chapter VII).

c. Mission Time Line Analysis

An MTLA is a graphical analysis which relates the sequence of tasks to be performed by the operator to a real time basis [Matheny, et al., 1970]. The purpose of an MTLA as used in the current study is to identify those performance measurement points within a man-machine system where standards of accuracy and time may be applied in the evaluation process. Essentially a bar chart, an MTLA for the navigation-to-TP segment of the radar navigation maneuver is presented as Appendix D. The time of execution for each subtask was extracted from estimated completion times on the task analysis record form (Appendix C). Time was estimated with the assumption that sensing conditions were good and the B/N was highly skilled. Darkly shaded time lines represent tasks that demand full mental attention whereas shaded time bars represent "monitoring" tasks or "troubleshooting" tasks that may not have to be executed.

The MTLA is a performance measurement source for both the identification of critical subtasks and the use of time to perform as a measure of skilled behavior. Thus, the MTLA was utilized to identify candidate performance measures as found in the "performance measure metric" column of the task analysis record form (Appendix C) that were later used for the final candidate measure set as will be described in Chapter VII (Table XI).

4. <u>B/N Skills and Knowledge</u>

This section will relate current skill acquisition principles to performance measurement, and present a skill acquisition model of the relationship between skill acquisition, the task, performance measurement, and performance evaluation. A great deal of discussion about the concept of skill, how skill is attained, and how skill acquisition is measured can be found in the literature from such diverse areas as private industry, control theory, information processing, and education [Bilodeau, 1966 and 1969; Jones, 1970; Welford, 1971; Hulin and Alvares, 1971 and 1971; Singleton, 1971; Leshowitz, et al., 1974; Shipley, 1976; Welford, 1976]. Despite the global interest in skill, this discussion will be limited to aircrew skill acquisition and the measurement of that skill.

a. Definition of Skill

Skill may be defined as the ability to perform given tasks successfully or competently in relation to specified standards [Cureton, 1951; Senders, 1974; Smit, 1976; Prophet, 1978]. A more precise definition of skill is offered by Connelly, et al. [1974]:

> The ability to use knowledge to perform manual operations in the achievement of a specific task objective in a manner which provides for the elimination of irrelevant action and erroneous response. This conceptualization exists only in conjunction with an individual task and is reflected in the quality with which this task is performed.

Most definitions of skill rely on the fundamental concept that the use of capacities efficiently and effectively as the result of experience and practice would generally characterize skill [Welford, 1976]. Indeed, the concept of skill cannot be well defined due to the diversity of its nature and remains more or less an amorphous quantity, best described by its characteristics [Singleton, 1971]:

1. It is continuous, there is always an extensive overlap and interaction. Even in principle, it cannot be analyzed by separation into discrete units along either space or time axes.

2. It involves all the stages of information processing identifiable in the organism, basically inputs, processing and outputs.

3. It is learned and therefore highly variable within and between individuals.

4. There is a purpose, objective, or goal providing meaning to the activity.

b. Skill Acquisition

The development of skill, as previously discussed, is due mainly to the effects of practice and experience on the use of basic capacities. Therefore, the acquisition of skill appears to result from learning and seems to improve the efficiency and effectiveness of underlying basic capacities [Welford, 1976]. Singleton [1971] advanced the idea that skill develops by selectivity and by the integration of activities. For most skill development theories, it is generally agreed that as learning a new task takes place, operators learn a basic strategy in performing the task, that in effect becomes an increasingly skilled template with the qualities of organizational and efficiency of operation [Engler, et al., 1980]. Depending on the task, the level of skill required to perform the task is universally measured with the property of variability. Bowen, et al. [1966] found considerable variability as measured by a lack of consistency for all skill levels of pilots performing tasks in an OFT, even pilots with substantial flight experience. It thus appears that as an operator begins to learn a new task, his control strategy in performing the task is highly inefficient, resulting in a large variability of actions. As skill development progresses, his control strategy becomes highly efficient and effective, resulting in what should be smaller variability of actions.

Three phases of skill development have been hypothesized in earlier research by Fitts [1962] and discussed in terms of aircrew skill acquisition by Smode, et al. [1962] and Prophet [1976]. The stages of skill acquisition are discussed below.

(1) Early Skill Development. In this phase the student seeks to develop a cognitive structure of the task in the form of discriminating the task purpose, ascertaining standards of task performance, and interpreting performance information feedback. Actions tend to be slow and deliberate, and depend a great deal on concentrated attention and effort in performing the task.

(2) Intermediate Skill Development. After learning the task purpose and experiencing some practice at the task, the student begins to organize his control strategy by becoming more efficient in responding to signals displayed to him. Perceptual and response patterns become fixed with less reliance placed on verbal mediation of response integration by the student.

(3) Advanced Skill Development. This phase represents the higher level of skill acquisition, where performance becomes more resistant to stress and activities are performed concurrently. The rate to acquire this stage through practice is different for each individual, as practice on any complex task generally spreads individuals out into stable but different skill levels [Jones, 1970]. Navigating an A-6E CAINS aircraft falls into this category of complex tasks. This stage is characterized by the individual performing in an automated manner requiring little conscious awareness and little allocation of mental effort [Norman, [1976].

c. Measurement of Skill Acquisition

Figure 3 is a model developed by the author to illustrate the relationship among B/N skill acquisition, the radar navigation task, and performance measurement and evaluation. An understanding of the model depends heavily upon concepts defined and discussed in Chapter IV and in the early part of this chapter. This model will be used for the current discussion of skill acquisition measurement.

The actual measurement of skill acquisition through its various stages has received little practical attention and research, most likely due to the complexity of the subject and the difficulty involved in accurately assessing human performance as system complexity increases [Glaser and Klaus, 1966; Vreuls and Wooldridge, 1977; Kelley and Wargo,

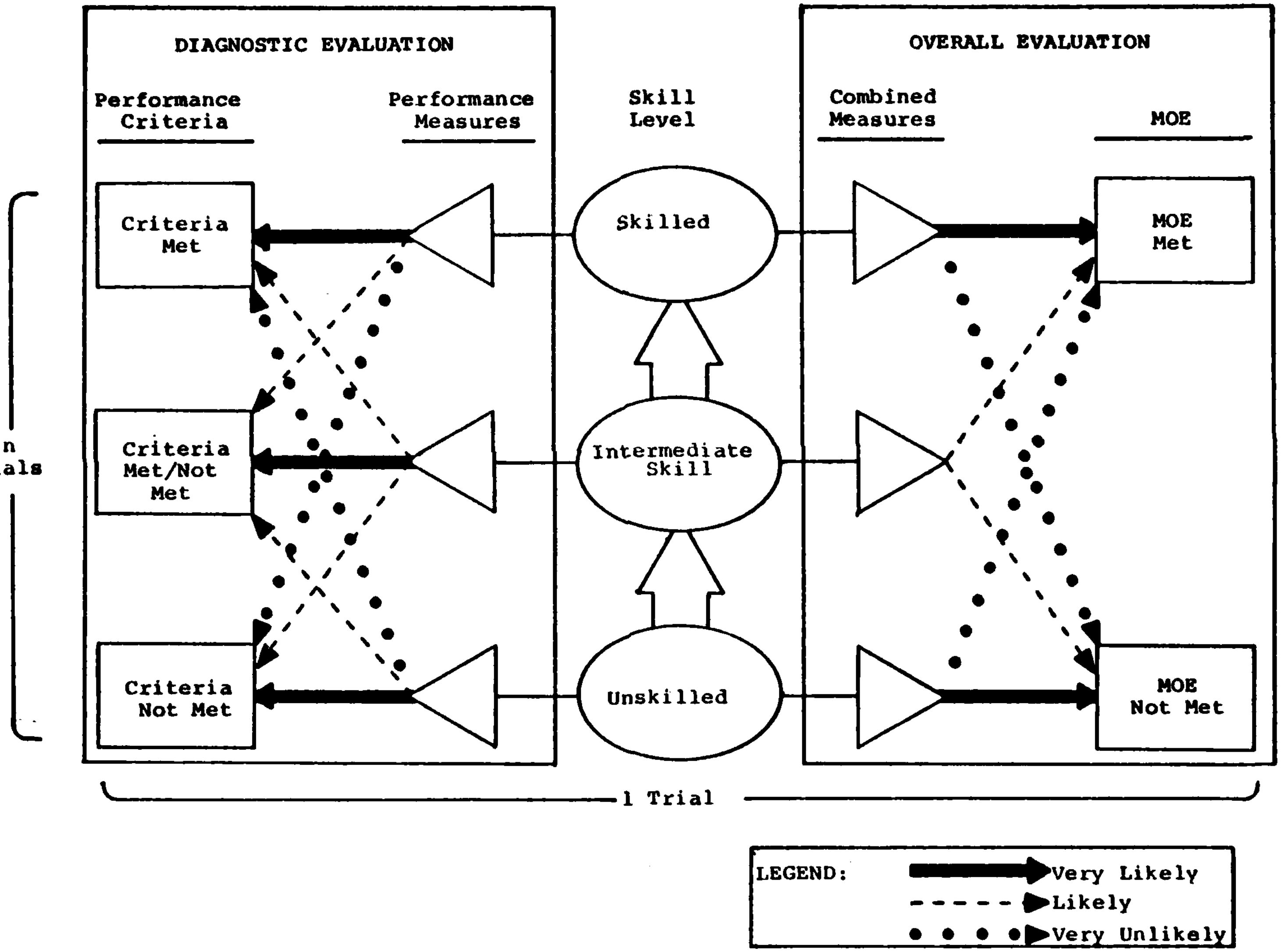


Figure 3. Skill Acquisition and Performance Evaluation Model.

1968; Senders, 1974]. As shown in Figure 3, through the process of learning and practice of the task over several trials, the student (represented by the oval shapes in the center column) should progress from the early or unskilled state through the intermediate stage and into the skilled or "proficient" stage where he is "trained." Over the course of one task trial, the most that can be accomplished is to obtain objective performance measures and combined measures, and to obtain a subjective opinion of the skill level from the one individual well qualified and skilled in performing the task: an instructor. Once this "indirect" measurement takes place, a comparison is made between the objective and subjective measures and the predefined performance criteria or MOEs. It is from this comparison that the student's skill level is finally evaluated, with severe limitations imposed due to the measurement over one trial. Skill acquisition through the three stages of development occurs not only at different rates, but over the course of several trials. This fact would lend support to a measurement system that indirectly measured skill development over one trial and used historical records of performance to measure skill development over several task trials. The model shows highly likely, likely, and very unlikely evaluation results by assuming that both criteria and measures are valid, reliable, and specific to the purpose of measuring skill acquisition.

Early conceptions of measuring aircrew skill were discussed by Smode, et al. [1962] and Angell, et al. [1964]. Both structured tasks or skills into a hierarchical model and theorized what types of measures (e.g., time, accuracy, frequency, etc.) would be appropriate for a particular level of task or skill. The former study also discussed measurement at the three stages of skill acquisition: (1) Measurement at the first stage should be concerned with knowledge and task familiarity as well as distinctions between task relevant and task irrelevant information and cues, and a differentiation between in-tolerance and out-of-tolerance conditions; (2) the intermediate stage has measurement concerned with procedure learning, the identification of action stimuli, and the performance of manipulative activities; and (3) measurement of highly developed procedural, perceptual-discriminative motor and concept-using skills and the integration of these combinations into more complex units of performance is of concern.

An experiment by Ryack and Krendel [1963] based on research by Krendel and Bloom [1963] measured highly skilled pilots performing a tracking task using a laboratory apparatus. The measurement was based on a theory that a highly skilled pilot displays consistency of system performance, is highly adaptable to changing dynamic requirements, and performs the task with least effort. The transfer of the measurement of these three conceptualizations of high pilot skill from the laboratory to an actual aircraft was not demonstrated.

Later conceptions of skill acquisition measurement were advanced by Welford [1971 and 1976] who proposed that as practice on a task increased, the speed of performance on that task as measured by time would fall exponentially. Haygood and Leshowitz [1974] proposed using an information processing model to measure flying skill acquisition. Bittner [1979] evaluated three methods for assessing "differential stability": (1) graphical analysis, (2) early versus late correlational Analysis of Variance (ANOVA), and (3) Lawley Test of Correlational Equality. That study recommended graphical analysis as a method of first choice.

Three recent experiments regarding measurement of aircrew skill acquisition are noteworthy. Vreuls, et al. [1974] used multiple discriminant and canonical correlation analyses to discriminate between different levels of skill using four pilots in an F-4E configured simulator. Using six pilots in a UH-1B (helicopter) simulator, Murphy [1976] investigated individual differences in pilot performance by measuring both man-machine system outputs and pilot control outputs during an instrument approach and landing. This study concluded that performance differences may be attributed to crewmember differences in cognitive styles, information processing abilities, or experience. Pierce, et al. [1979] concentrated on procedures to assess cognitive skills through the use of behavioral data on eight pilots performing F-4 aircraft pop-up maneuvers. The primary measurement instrument was an instructor using subjective ratings that were validated by comparison to actual bomb scores.

From this previous research and the discussion of the skill acquisition measurement model, it becomes readily apparent that measurement of B/N skill acquisition in the A-6E WST during radar navigation will require both an analytical foundation, as described in this thesis, and empirical validation that would result from implementation of the proposed measurement system. This section is concluded with six recommendations for the procedure of skill appraisal, as discussed by Singleton [1971]:

(1) Discuss the skilled activity almost _ad nauseam_ with the individuals who practice it and with those to whom and for whom they are responsible. It is not enough to pop in at intervals, the investigator must spend whole shifts and weeks with the practitioners to absorb the operational climate.

(2) Try to make this verbal communication more precise by using protocol techniques, critical incident techniques, good/poor contrast techniques, and so on.

(3) Observe the development of the skill in trainees and by analysis of what goes on in the formal and informal training procedures and in professional assessment. Make due allowance for history, tradition, technological change, and so on.

(4) Structure the activity. Identify the dimensions of the percepts, the decision making, the strategies of action and the overt activities, and try to provide scales of measurement along each dimension.

(5) Check as many conclusions as possible by direct observation, performance measurement, and by experiment.

(6) Implement the conclusions and provide techniques for assessing the limitations and successes of the innovations.

VI. A-6E WST-PERFORMANCE MEASUREMENT SYSTEM

The A-6E WST, Device 2F114, is designed to provide full mission capability for pilot transition training, B/N transition training, integrated crew training, and maintenance of flight and weapon system proficiency in the A-6E Intruder aircraft. The WST will be used to train Navy/Marine flight crew members in all A-6E procedures - ground handling, normal and emergency flight modes, communications, navigation and cross-country missions, tactics, and crew coordination [Read, 1974]. Inherent in these design and training requirements is the necessity for the measurement of aircrew performance and the subsequent measurement of improved performance after training has occurred; a necessary goal for any training simulator [Knoop, 1968]. This section will discuss the general characteristics of the WST together with current performance measurement capabilities, generic performance measurement systems (PMS) for simulators, and current performance measurement and evaluation practices for student B/Ns in the WST.

A. WST CHARACTERISTICS

1. General Description

The trainer system consists of the following elements: Trainee Station, Instructor Station, Simulation Area, and Mechanical Devices Room (see Figure 4). The Trainee Station

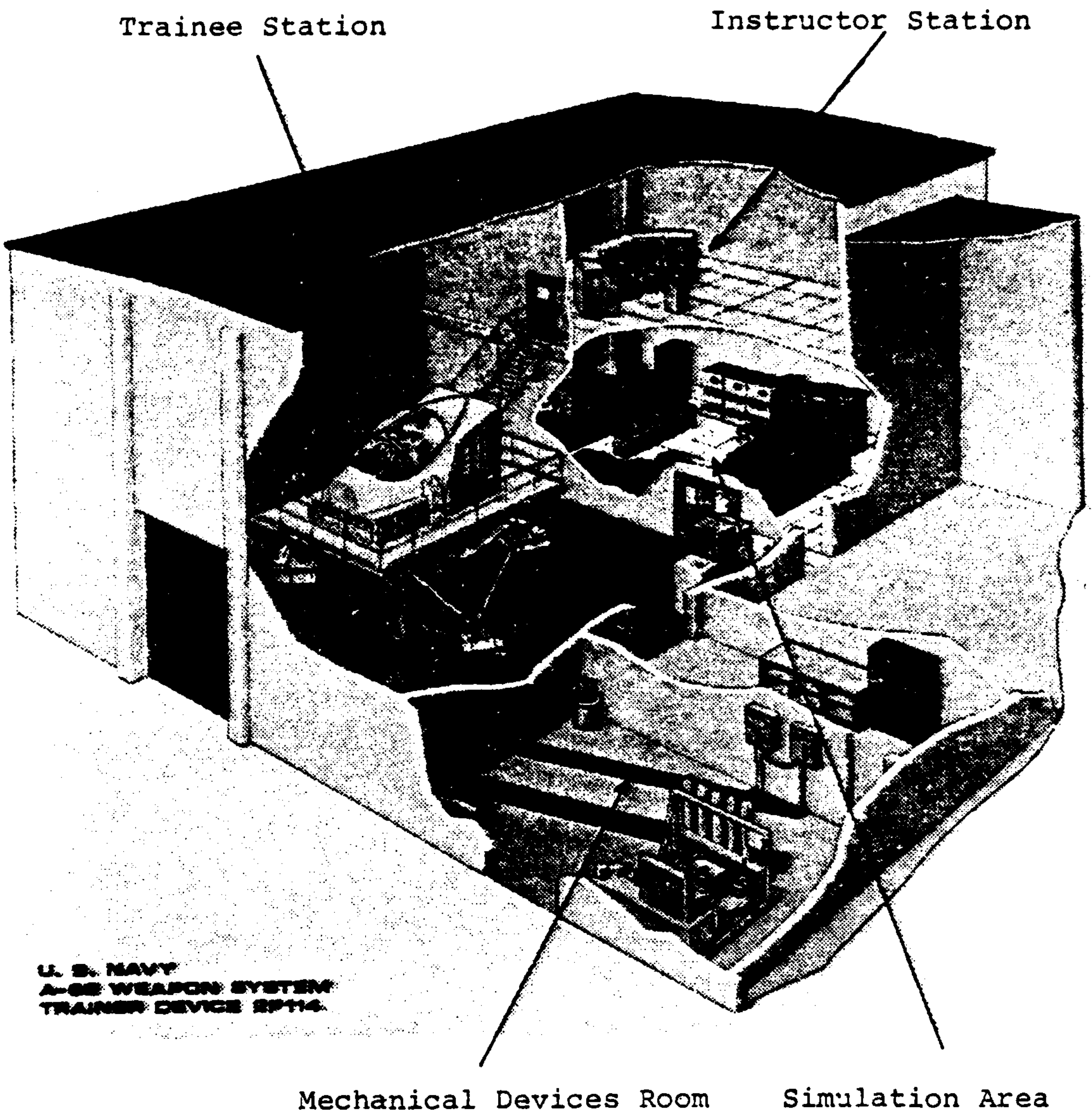


Figure 4. Device 2F114, A-6E WST.

is an exact replica of the A-6E CAINS cockpit and is mounted on a six degree-of-freedom motion base to give realistic motion cues. Sound cues, environmental controls, and controls with natural "feel" increase the similarity between WST and aircraft cockpits. Normal and emergency flight configurations are simulated, together with all modes of weapon system operation.

The Instructor Station area consists of a wrap-around console that can accommodate two principal instructors and two assistant instructors. Controls and displays are utilized by the instructors to: (1) set up and control the training problem, (2) introduce malfunctions and failures, (3) monitor trainee actions and responses to malfunctions, and (4) evaluate trainee performance. Four interactive CRT displays for alphanumeric and graphic presentations, together with repeater displays of the VDI, direct view radar indication (DVRI), and electronic countermeasures (ECM) found in the aircraft are available to the instructors for use in training.

The Simulation Area contains the computers necessary for simulation of the A-6E CAINS aircraft and its missions. Four real-time minicomputers are utilized, two for flight, one for tactics, and one for the Digital Radar Land Mass Simulation (DRLMS). Magnetic tape units, teletypewriter printers, digital conversion equipment, and the DRLMS are also contained in this area. The DRLMS is designed to simulate landmass radar return for the AN/APQ-156 radar, which is a major component of the A-6 navigation/weapon system and is used by the B/N for the tasks of radar scope interpretation and target location.

The Mechanical Devices Room contains hydraulic and power equipment for positioning of the Trainee Station motion system. Also provided from this area is compressed air needed for g-suit and environmental control requirements.

2. Performance Evaluation System

A comprehensive list of system features and characteristics is beyond the scope of the present effort, but can be found in the Grumman Aerospace Corporation Final Configuration or Criteria Reports for the A-6E WST [Blum, et al., 1977, 1977, and 1977; Rinsky, 1977]. Those features which are of interest to performance evaluation in the WST are shown in Figure 5 and discussed below:

a. Program Mission Modes are provided for up to ten missions. In this mode, the computer system automatically generates and sequences mission profiles. During this mode, the instructor station monitor displays a listing of the mission leg number, maneuver to be performed, mission leg end, parameters to be monitored, and remarks in order of occurrence. Programmed missions are activated by controls from the instructor station.

b. Program Mission Mode Critiquing is provided by instructor selection of up to six parameters for monitoring on each leg of the program mission. The system computes the

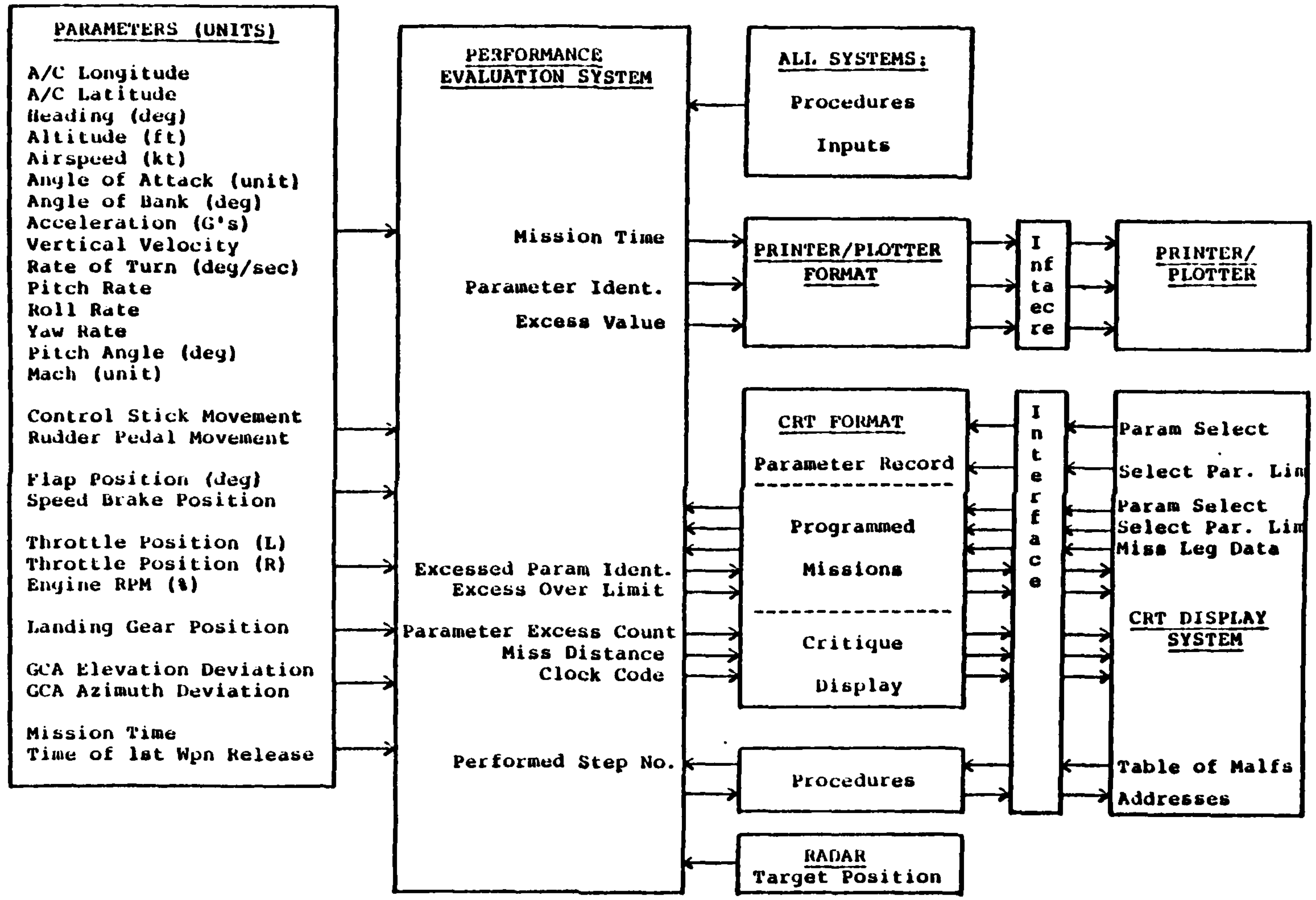


Figure 5. A-6E WST Performance Evaluation System.

difference between the parameter and a tolerance value for the parameter preselected by the instructor. When the tolerance is exceeded for a selected parameter, the exceeded amount is displayed to the instructor and a printout record is provided at rates of every 5 seconds, 10 seconds, 15 seconds, 30 seconds, 1 minute, and 2 minutes on a printer/plotter. Available parameters are shown in the left-hand column of Figure 5.

c. Procedure Monitor Display is a mode in which all steps of up to any two procedures, normal or emergency, appear automatically on the instructor's display system when called up by the instructor. The text for the procedures and malfunctions is a listing of the steps to be performed by the trainee and an indication of the elapsed time required by the trainee to complete the procedure. This mode is also available during the Programmed Mission Mode.

d. Parameter Recording is available on all modes of WST operation. A minimum of six parameters may be simultaneously recorded on a continuous basis as a function of time, and compared to preselected tolerance values as discussed in (b) above. The parameter record mathematical model program is available in the flight simulation computer.

e. CRITIQUE Mode calculates miss distance and clock code of trainee delivered weapons on an instructor-designated target by the use of a Scoring Math Model. Bomb circular probable error (CEP) is calculated using available functions and parameters at time of release. Missile releases are scored as "hit" or "miss" based on computed comparisons to a respective missile envelope. A CRITIQUE display for a permanent record of the results is available.

f. Event Recording is provided where up to thirty events selected by the instructor are monitored by the Performance Evaluation System. After the event is selected for recording by the instructor, a printout is initiated which contains a statement of the event, other parameter values, and the time of occurrence. Table VIII is a listing of available events for recording along with other recorded parameters.

g. Audio voice recording with a time mark and rapid recall function permits the instructor to access desired portions of trainee headset radio during and after a training mission. All pertinent communications can be recorded for up to 2.5 hours.

h. Navigational computations, display drive signals and positional readouts were designed to be within 0.1 nautical mile of true position based on ground speed and true course.

i. A Versatec electrostatic Printer Plotter unit is furnished at the instructor console area and has the capability of simultaneously printing and plotting parameter recordings, Program Mission parameter recordings, and event recordings during Free Flight, Program Mission, and Demonstration Maneuver/Trainee mode training missions. This unit can also

TABLE VIII. A-6E WST EVENT RECORDING SELECTION

	EVENT	OTHER EVENT PARAMETERS
1.	Airborne	-
2.	Gear-up and locked	IAS
3.	Flaps/Slats-up	IAS
4.	Stability Augmentation Engaged	ALT
5.	Search Radar Switch Stby/on	-
6.	Doppler Radar Switch Stby/on	-
7.	Chaff emitted	-
8.	Isolation Valve Switch FH/Land	-
9.	Fuel Dump (wing or FUS)-on/secured	-
10.	Tank Pressure Switch-on/off	-
11.	Present Position Correct Button-Depressed	-
12.	Computer - on/off	-
13.	Computer Error Lite Lit	-
14.	Master Caution Lite Lit	-
15.	ALQ-126 Switch - Rec/Repeat	-
16.	Reselect Lite Flashing/Steady	-
17.	Master Arm - on/off	-
18.	Attack; Step In to/Out of	-
19.	Bomb Release	-
20.	Commit Trigger - Depressed	-
21.	AZ Range Switch - on/off	-
22.	Velocity Correct Switch - Memory/Off Save	-
23.	Track-While-Scan; on	-
24.	Computer - Out of Attack	-
25.	Throttle(s) below 75 percent	-
26.	Gear Handle Down	IAS
27.	Flap/Slat Lever - 30/40 degrees	IAS
28.	Touchdown	IAS, AOA
29.	Ram Air Turbine - in/out	-
30.	Designate - on/off	Slant Range

Source: Blum, et al. [1977]

print out any display type designated by the instructor with a maximum of 20 printouts possible during any 2.5-hour mission.

j. The tactics computer exercises master control of the system and includes functional control of the attack navigation system, system displays, weapons release system, inflight refueling system, ECM, threats, magnetic variations, Programmed Missions, dynamic replay, malfunctions, instructor display system, malfunction control, displays, instructor flight control, demonstration maneuvers, CRITIQUE mode, and others. This computer was installed with future hardware and software growth for input-output, memory core, and computation as a design specification.

The performance measurement capability of the A-6E WST appears to have an impressive objective measurement capability. The hardware and software computer system was designed with objective performance measurement in mind, although no definite model or technique was provided by the designers for evaluating B/N skill acquisition during a radar navigation mission. The foundation has been laid for objective measurement; all that remains is building a sound performance measurement structure based on principles and models that have been examined and evaluated thus far.

B. PERFORMANCE MEASUREMENT SYSTEMS

Measuring aircrew performance can be viewed as a system within itself. Every system consists of an assemblage or combination of objects or parts forming a complex or unitary whole with definable characteristics and a common purpose. The purpose of the performance measurement system (PMS) examined here will be to provide FRS instructors and training managers with valid, reliable, and objective information needed to guide decisions about trainee skill acquisition. The PMS for the similator has definable components, functions, inputs, outputs, communication links, and procedures that all interact to form a system that may or may not be efficiently designed or implemented. Criteria for PMS selection may outweigh some desirable system characteristics as well as the optimal allocation of functions to a particular component. After a PMS has been analyzed, functions allocated, and system criteria selected, implementation of the PMS within the operational environment may impose further constraints that cause redesign of the system. The interactions of PMS analysis, functional allocation, criteria, and implementation are discussed below.

1. Systems Analysis

Since the purpose of the PMS being discussed is to provide information about student skill level to training personnel for accurate training control decision-making, this system can be viewed from an information-processing approach. Information in the form of data is sensed and collected, recorded and processed, and presented in an output form that is useful for performance evaluation purposes. These functions of the system are interdependent and may be served by the same component. Major components of the PMS are instructors and computers with data storage capability. Discussion of the PMS analysis follows.

a. Data Sensing and Acquisition

Performance must be observed to be sensed and collected. Performance measurement considerations in data collection include but are not limited to: (1) mission purpose, (2) flight regime, (3) maneuver performed, (4) tasks, (5) skills required, (6) operator physiological output measures, (7) aircraft measures, (8) aircrew-aircraft system output measures, (9) mission results, (10) flight management, (11) procedural control, (12) aircraft systems management, (13) operator motivation, and (14) historical data.

Sensing and collecting performance information in a simulator can be accomplished by: (1) mechanical and electronic devices including digital computers and (2) direct human observation [Angell, et al., 1964]. The first category is usually referred to as "automated" measurement devices, and may include video/photo recorders, audio/digital recorders, timers and counters, graphic recorders, and plotters [Smode, et al., 1962; Angell, et al., 1964; Obermayer, et al., 1974; Hagin, et al., 1977]. Direct human observation may or may not be standardized by preplanned performance checklists or instructions.

Video/photo recorders provide permanent records of performance and are suitable for instructors to use in observing performance more objectively because of playback features. Audio/digital recorders involve recording of communications and direct measurement conversion to digital form by a computer for selected observable parameters. Audio recordings may be utilized as a more objective measurement for instructor use but currently have data conversion limitations. Digital recording of discrete and continuous measures from all levels of aircrew-aircraft system performance has been demonstrated in both simulators and in actual aircraft flight over the past twenty years [Wierwille and Williges, 1978; Mixon and Moroney, 1981]. Timers and counters are suitable as auxiliary components to digital computer measurement for both time and frequency performance measures [Angell, et al., 1964]. Graphic recorders are electromechanical in operation and provide continuous records of event states and magnitudes along a time continuum. Graphic recorders are usually either classified as event (discrete performance) or continuous (magnitude of continuous variable) [Angell, et al., 1964]. Plotters display information in Cartesian or rectangular coordinates, and are useful for both performance data collection and output.

Charles [1978] thoroughly studied the role of the instructor in a simulator and determined one function of the instructor was to monitor performance in the form of student procedures, techniques, skill level, and simulator performance. Indeed, the direct observation of student performance may sometimes be the sole source of valuable performance measurement, especially when unexpected events occur during a simulator mission [Smode, et al., 1962]. As previously mentioned, video recording with playback capability improves the human observation data collection method. Usually, performance measurements resulting from this technique must be converted for digital computer use in the data playback and processing stage.

b. Data Processing and Analysis

Once performance data are sensed and collected by mechanical or human means, some conversion is usually required to make the raw data more useful for the system purpose, i.e., to provide information for accurate performance evaluation. Usually all data are converted to a digital format appropriate to the general purpose computer. It is in this stage where computers and peripheral equipment such as input/output devices, memory core units, and magnetic tape drives are extremely accurate, efficient and cost-effective as compared to human processing of data, although some data types may not be convertible to a digital format and must be carried to the system output stage in raw form. In this stage, usually video recordings are reviewed by the instructor to increase the objectivity of his direct human observation of performance. For the interested reader, Obermayer and Vreuls [1974] present a more detailed account of data playback and processing components and interactions.

c. Data Presentation

After data analysis, the data will be available as output measures for the evaluation process. The output format may be numerical, graphical, audio, visual, or some other form. Since the evaluation process involves the comparison of performance data to standards or criteria, some of the performance data may be utilized as criteria for subsequent evaluation use. Most likely, the data output will be typical measures of time, accuracy, and frequency for various task levels. Some measures may be combined in the processing stage and used as output data for comparisons to established MOEs.

2. Allocation of Functions

One result of the systems analysis of the simulator PMS was to identify functions that have to be performed. Given there may be an option as to whether any particular function should be allocated to the human or a machine, some knowledge of the relative capabilities of humans and machines would be useful for determining the allocation of functions. Some relative capabilities among mechanical devices and human observers were discussed in the previous section but more detail is necessary. Using the results of McCormick [1976], Buckhout and Cotterman [1963], Obermayer, et al. [1974], Angell, et al. [1964], and Coburn [1973], the capabilities of humans and machines for the purpose of performance measurement in the simulator are presented in Table IX.

TABLE IX: HUMAN AND MACHINE CAPABILITIES AND LIMITATIONS

<u>HUMAN CAPABILITIES</u>

Detect stimuli against background of high noise (CRT).

Recognize patterns of complex stimuli (DVRI).

Sense and respond to unexpected events.

Store large amounts of diverse information for long periods.

Retrieve information from storage (with low reliability).

Draw upon experience in making decisions.

Reason inductively, generalizing from observations.

Apply principles to solutions of varied problems.

Make subjective estimates and judgements.

Develop entirely new solutions.

Select only most important events for sensing inputs.

Acquire and record information incidental to primary mission.

High tolerance for ambiguity, uncertainty, and vagueness.

Highly flexible in terms of task performance.

Performance degrades gradually and gracefully.

Override own actions should need arise.

Uses machines in spite of design failures or for a different task.

Modify performance as a function of experience.

<u>MACHINE CAPABILITIES</u>

Sense stimuli beyond man's range of sensitivity.

Apply deductive reasoning when classes are specified.

Monitor for prespecified frequent and infrequent events.

Store coded information quickly and in quantity.

Retrieve coded information quickly and accurately.

TABLE IX (Continued)

MACHINE CAPABILITIES (Continued)
Process quantitative information.
Make rapid, consistent, and repetitive responses.
Perform repetitive and concurrent activities reliably.
Maintain performance over time.
Count or measure physical quantities.
Transfer function is known.
Data coding, amplification, and transformation tasks.
Large channel capacity.
Not influenced by social and physiological factors.
HUMAN LIMITATIONS
Sense stimuli within a limited range.
Poor monitoring capability for activities.
Mathematical computations are poor.
Cannot retrieve large amounts of information rapidly and reliably.
Cannot reliably perform repetitive acts.
Cannot respond rapidly and consistently to stimuli.
Cannot perform work continuously over long periods.
Requires time to train for measurement and evaluation.
Expectation set leads to "see what he expects to see."
Requires review time for decisions based on memory.
Does not always follow an optimum strategy.
Short-term memory for factual material.
Not suited for data coding, amplification, or transformation.
Performance degraded by fatigue, boredom and anxiety.

TABLE IX (Continued)

HUMAN LIMITATIONS (Continued)
Cannot perform simultaneous tasks for long periods.
Channel capacity limited.
Dependent upon social environment.
MACHINE LIMITATIONS
Cannot adapt to unexpected, unprogrammed events.
Cannot learn or modify behavior based on experience.
Cannot "reason" or exercise judgement.
Uncoded information useless.
Inflexible.
Requires stringent environmental control (computers).
Cannot predict events in unusual situations.
Performance degraded by wearing out or lack of calibration.
Limited perceptual constancy and are expensive.
Non-portable.
Long-term memory capability is expensive.
Generally fail all at once.
Little capacity for inductive reasoning or generalization.

When fully exploited with no other limitations imposed, these capabilities and limitations of the human and machine define what might be described as an "optimal" performance measurement system from the engineering standpoint. As Knoop and Welde [1973] observed, a performance measurement system should "capitalize on the advantages of an automated, objective system and yet retain some of the unique capabilities afforded by the human evaluator." For each task which is to be measured and evaluated, a decision must be made as to whether it would be more efficient for the man or the machine to measure or evaluate performance on that task [Buckhout and Cotterman [1963].

3. <u>System Criteria</u>

In addition to examining human and machine capabilities and limitations for the functions and components of a performance measurement system, other factors with potential impact on system design and implementation must be identified, analyzed, and weighed for importance. Choosing a system solely by human-machine advantages does not take into account other apparently extrinsic influences that may turn out to be deciding factors. The following listing of system criteria for performance measurement systems was gleaned from research by Obermayer and Vreuls [1974], Buckhout and Cotterman [1963], Demaree and Matheny [1965], Farrell [1974], and Carter [1977]:

a. Conflicts of system purpose may exist. The PMS is required to provide objective, reliable, and valid information for decision-making purposes and to also identify changes in student skill level. A component may be the most objective choice available for the first goal but inadequate for the second. An alternative component or function may be identified to do both satisfactorily.

b. Data should be provided in a useful form for evaluation purposes.

c. Data collection, processing, and presentation must be timely enough to enhance the training process in the form of knowledge of results.

d. Costs to modify or supplement equipment and software must be weighed against the utility of the information derived.

e. *Data distortion* must be controlled for accurate and reliable results.

f. Minimum interference with the training process should occur with the measurement system having an inconspicuous role requiring little or no attention from the student or instructor.

g. Social impacts of any system may have adverse effects on morale or personnel involvement. If an instructor perceives that automated performance measurement is a replacement to his traditional role as an evaluator, the effectiveness of the measurement system will be greatly reduced.

h. Economic and political constraints may affect system design. The ideal measurement device may not be recommended for procurement by higher authority, while some selection of components is based on available equipment at time of procurement.

i. Other factors such as size, weight, safety, ease of use and reliability should also be considered.

Obviously, system criteria should be used in the sense that selecting components and functions for a performance measurement system would maximize those criteria that are advantageous and minimize those aspects that are not optimum for the system purpose. These criteria must be taken into account during allocation of functions for the performance measurement system, and must be weighed at least qualitatively if not in a quantitative sense for overall contribution to the final system configuration.

4. System Implementation

Once the objectives of the performance measurement system are identified and the allocation of functions and system criteria are applied, the system model then requires a deliberate implementation procedure if it is to produce meaningful results and have utility to the end user. Waag, et al. [1975] identified four phases of development for implementation of the measurement system in the simulator:

(1) Definition of criterion objective in terms of a candidate set of simulator parameters.

(2) Evaluation of the proposed set of measures for the purpose of validation and simplification.

(3) Specification of criterion performance by requiring experienced instructor aircrew to fly the maneuver in question.

(4) Collection of normative data using students as they progress through the training program.

When using automatic or human measurement components within the performance measurement system, other implementation considerations should apply. These are discussed below.

a. Automatic Measurement Considerations

In simulator environments, the organization of the software will be the key to successful implementation of flight training measurement systems [Vreuls and Obermayer, 1971]. Extensive research into programming techniques for the automatic monitoring of human performance in the flight regime has been accomplished [Knoop, 1966 and 1968; Vreuls and Obermayer, 1974; Vreuls, et al., 1973, 1974, and 1975]. Knoop [1968] examined some of the prerequisites for automatically monitoring and evaluating human performance:

(1) Knowledge is required of which performance variables are important in evaluating an operator's proficiency.

(2) Knowledge is required of how these variables should be related for optimal performance.

(3) A digital computer program is required which compares actual relationships among these variables during performance with those required for optimal performance to evaluate operator proficiency.

These prerequisites point out the need for careful front-end analysis of the system in terms of performance measures and criteria, and the complex problem involved of programming this analysis for automatic measurement and evaluation.

b. Human Measurement Considerations

Smode [1962] provides some rules for enhancing the validity and reliability of resulting measurement where human observers are employed in data collection:

(1) Provide standardized checklists that specify what to observe, when, and how often observations are recorded.

(2) Train observers for the measurement process to insure full understanding of how to make the observations.

(3) Provide data collection sheets that conveniently indicate what is to be observed and the sequence of observation.

(4) Avoid overloading the observer with too much simultaneous observation and recording.

(5) Data collection forms should have notation or symbology for recording observations, when feasible, and should result in a permanent record that can be easily transformed into a form for rapid analysis.

These guidelines still appear sensible today, with perhaps some additional information about the relationship between automatic measurement and the human observer being established and provided within the system.

C. CURRENT PERFORMANCE MEASUREMENT IN THE WST

Navy B/N replacement training is conducted by both the East Coast FRS, Attack Squadron Forty-Two (VA-42), and the West Coast FRS, Attack Squadron One Twenty Eight (VA-128). Each FRS has developed and maintains its own training program; however, these programs are similar in nature and utilize virtually the same performance measurement and evaluation techniques. Each training program is divided into specific phases designed to develop certain skills such as navigation, system operation, or attack procedures, and each uses the mediums of classroom lecture, simulator, or actual aircraft flight. A building-block approach to developing progressive knowledge of skills is utilized, including training missions in the WST. This study will focus on a Category One (CAT I) B/N student, where entry skills and knowledge for measurement in the WST are minimal.

Determination of the skill level of CAT I B/Ns performing simulated missions in the A-6E WST continues to be based on subjective judgements made by instructor B/Ns and pilots. A syllabus of progressively more difficult flights in the WST is part of the training curriculum. During or shortly after each flight, the instructors "grade" the student on tasks performed employing mostly personal criteria, based on experience and normative comparisons of the student's performance with other student performances. Table X is a compilation of tasks taken from B/N flight evaluation sheets for the VA-42 simulator curriculum for which B/N performance is graded on a scale using four categories: (1) unsatisfactory, (2) below average, (3) average, and (4) above average. A typical B/N flight evaluation sheet for the WST is shown in Figure 6. The four ratings listed above are then converted to a 4.0 scale for numerical analysis and student rankings.

During a personal visit by the author to the VA-42 A-6E WST in June 1980, subjective performance measurement

TABLE X: CURRENT B/N TASKS GRADED IN WST CURRICULUM

Aggressiveness
Aircraft familiarity
Aircraft system knowledge - NATOPS
Aircraft/system operations
Aircraft/system turnup
Aircraft/system utilization
ALE-39 awareness
ALQ-126 awareness
ALR-45/50 awareness
AMTI
Approach
Approach (TACAN, GCA, ASR)
Attack procedures/ACU knowledge
Attitude
Basic air work
BINGO procedures
Communications
Computer operation
Crew concept
Degraded aircraft/system operation
Degraded aircraft/system utilization
Degraded system CEP
Degraded system utilization
Departure
Departure procedures
ECM equipment knowledge
ECM tactics
Emergency procedures

TABLE X (Continued)

Flight briefing
Fuel management
Full system utilization
General attack CEP
Glideslope control
Headwork
HI Loft attack CEP
Impact accuracy
Knowledge of the cockpit
LABS attack CEP
Landing transition
Line-up corrections
Low level navigation
Marshall pattern
Mining procedures
NATOPS
Navigation procedures
NORDO procedures
Normal procedures
Planning/preparation
Point checks
Post landing/shutdown procedures
Post start
Prestart
Radar interpretation
Radar operation
S-1 pattern
S-3 pattern

TABLE X (Continued)

Shipboard procedures
SRTC utilization
Stall series
Start
Start/point checks
Straight path attack CEP
System shutdown
System turnup
Takeoff/departure
Target procedures
Targets (geographical turn point listed)
Turn point acquisition
UHF communications
Use of checklists
Use of the clock

B/N FLIGHT EVALUATION SHEET

>>> BCW01 <<<

REPLACEMENT: BUNO:________

INSTRUCTOR____________________ DISPOSITION CODE:________

FLIGHT TIME: TOTAL__________ NIGHT________ INST________

BRIEF TIME________ T/O TIME________ DEBRIEF________

	ITEM	UN	BA	A	AA
1.	DEPARTURE PROCEDURES........				
2.	MARSHALL PATTERN............				
3.	APPROACH....................				
4.	GLIDESLOPE CONTROL..........				
5.	LINE-UP CORRECTIONS.........				
6.	COMMUNICATIONS..............				
7.	NORDO PROCEDURES............				
8.	BINGO PROCEDURES............				
9.	EMERGENCY PROCEDURES........				
10.	HEADWORK....................				
11.	CREW CONCEPT................				
	TOTAL				

COMMENTS:

SIGNATURE________________________ DATE:__________

Figure 6. Typical B/N Flight Evaluation Sheet for WST

and evaluation was exclusively being conducted for training missions of students in the WST. The use of subjective performance measurement and evaluation was due mainly to the newness of the simulator and the traditional and acceptable role that subjective methods have played for the last three-quarters of a century across all aviation communities. Operational FRS personnel rarely have the time to carefully analyze and employ new performance measurement models and techniques or utilize new systems that are incorporated into a newly-delivered simulator. One purpose of this thesis is to eliminate the gap of carefully analyzing and evaluating performance measurement models for operational use.

Due to the exclusive use of subjective performance methods, standards of performance in the A-6 FRS are established analytically based on perceptions by the instructors on what constitutes "proficient" or "skilled" performance. Bombing and radar target identification (RTI) criteria are used internally, with RTI grading based on target difficulty and the replacement B/N's level of exposure and experience. Essentially, RTI criteria use a trinomial division of "hit the target," "in ball park," and "out of ball park" that has been defined well enough to be converted to a numerical grade. In addition, the replacement B/N must identify 75 percent of the assigned targets on a specific radar navigation "check flight" as a criterion for radar navigation skill.

The proposed model for measuring B/N performance during radar navigation in the WST, to be discussed in Chapter VII, incorporates the best qualities of both subjective and objective measurement, and uses the results to provide accurate and valid information for making decisions about student progress within the training process. Under the proposed model, criteria for successful performance will be established empirically for operational use by either the A-6 FRS or the A-6 community as a whole.

VII. RESULTS AND DISCUSSION

The purpose for designing a performance measurement system for the B/N during radar navigation in the A-6E WST was to provide objective, reliable, valid, and timely performance information for accurate decision-making to the training mission instructor and FRS training manager. This section presents the performance measurement system model developed from the previous analysis of related aircrew performance literature, generic performance measurement systems concepts, the B/N radar task analysis and MTLA, and the A-6E crew-system network model. The model developed is specific from the standpoint of identifying what to measure, when to measure, scaling, sampling frequency, criteria establishment, applicable transformations, observation method, current availability in the A-6E WST, and the accessibility of the measure if not currently available. The proposed model embodies: (1) the establishment of standards of performance for the candidate measure set by utilizing fleet-experienced and motivated A-6E aircrews performing well defined radar navigation maneuvers and segments, (2) techniques for reducing the candidate measures to a small and efficient set by statistical analysis, (3) evaluation methods which use the results from established performance standards and performance measurement of student B/Ns for decision analysis by the FRS instructor and training manager, and (4) some performance measurement informational displays which present diagnostic and overall evaluation results in a usable and efficient format.

A. CANDIDATE MEASURES FOR SKILL ACQUISITION

Using the candidate performance measure metrics derived from the B/N task analysis (Appendix C) and previous aircrew research (Table II), a composite list of candidate measures for B/N skill acquisition is presented as Table XI. Information is provided for each measure in terms of the method of measurement, measure segment, scaling, sampling rate, criteria establishment, transformations, availability in the A-6E WST, and accessibility in the A-6E WST (if not available). Each of these terms is defined below.

1. Method of Measurement

Either electronically (E), instructor observation (O), or both. The primary basis for the determination of the best method was both measure selection criteria (Chapter IV) and human and machine capabilities and limitations (Table IX).

2. Measure Segment

This is the period of time or segment of flight in which the measure should be observed. "ENTIRE LEG" defines the radar navigation segment from TP to TP, "MISSION" defines the segment from takeoff to landing, and "TP" defines within one minute of approaching the TP.

TABLE XI: CANDIDATE PERFORMANCE MEASURES

	MEASURE			
	Communication time (UHF and/or ICS) or total seconds of time spent communicating	Proportion of time spent communicating (UHF and/or ICS) as compared to total time period	Communication effectiveness (UHF and/or ICS)	Number of communications (UHF and/or ICS)
METHOD OF MEASUREMENT	E	E	O	E
MEASURE SEGMENT	Entire leg	Entire leg	Mission	Entire leg
SCALE	0 - 1000 sec	0 - 1	0/1	0 - 500
SAMPLING RATE	1/sec	1/leg	1/Mission	1/sec
CRITERIA	EMP	EMP	OPER	EMP
TRANSFORMATION	MIN, ME, MO, MAX	MIN, MO, MAX	PROPORTION	MIN, ME, MO, MAX
CURRENTLY AVAILABLE				
ACCESSIBLE	X	X	X	X

TABLE XI: Contd

	MEASURE			
	Total time to insert data into computer using keyboard (COMPTMODE SW from "ENTER" to "STEER")	Proportion of time spent entering data into computer	Number of times data entered into computer	Number of times system present position or altitude are read (DDU DATA SW thrown to "PRES POS")
METHOD OF MEASUREMENT	E	E	E	E
MEASURE SEGMENT	Entire leg	Entire leg	Entire leg	Entire leg
SCALE	0 - 1000 sec	0 - 1	0 - 500	0 - 500
SAMPLING RATE	1/sec	1/leg	1/sec	1/sec
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	ACC, MIN, ME, MO, MAX	MIN, MO, MAX	MIN, ME, MO, MAX	FREQ, ME, MO
CURRENTLY AVAILABLE				
ACCESSIBLE	X	X	X	X

TABLE XI: Contd

	MEASURE			
	Number of Lat/Long Data Insertions into Computer (POS Action Key)	Number of Times Data Entry is Verified for Accuracy (COMPTMODE SW to "STEER" Followed by DDU DATA SW to "ON CALL")	Total Time Spent Tuning Radar Scope (All Radar Control Panel Components)	Proportion of Time Spent Tuning Radar Scope
METHOD OF MEASUREMENT	E	E	E	E
MEASURE SEGMENT	Entire leg	Entire leg	Entire leg	Entire leg
SCALE	0/1	0 - 500	0 - 1000 sec	0 - 1
SAMPLING RATE	1/sec	1/sec	1/sec	1/leg
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	FREQ, PROPORTION	FREQ, ME, PROPORTION	TIME, MIN, ME, MO, MAX	MIN, MO, MAX
CURRENTLY AVAILABLE				
ACCESSIBLE	X	X	X	X

TABLE XI: Contd

	MEASURE			
	Outbound Heading Accuracy (HSI Pull to Set Knob)	Heading Error at TP Passage (True Inbound Heading)	Number of Turns to Assigned Heading	TP Passage Identification (DVRI Bug Movement to 180° Relative Position
METHOD OF MEASUREMENT	E	E	E	E
MEASURE SEGMENT	TP	TP	Entire leg	TP
SCALE	0 - 360 deg, 0/1	0 - 360 deg	0 - 500	0.1
SAMPLING RATE	1/sec	1/sec	1/sec	1/sec
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	TIME, FREQ, ACC, ME, PROPORTION	ACC, ME, MO	FREQ, ME	REACTION TIME, FREQ
CURRENTLY AVAILABLE		X	X	
ACCESSIBLE	X			X

TABLE XI: Contd

	MEASURE			
	Clock Activation Error at TP Passage (Leg Time only)	System Steering Error Recognition Time (from Onset of Error)	System Present Position Error Recognition Time (from Onset of Error)	System Steering Trouble-shooting
METHOD OF MEASUREMENT	E	E, O	E, O	O
MEASURE SEGMENT	TP	Entire leg	Entire leg	Mission
SCALE	0 - 200 sec, 0/1	0 - 1000 sec, 0/1	0 - 1000 sec, 0/1	0/1
SAMPLING RATE	1/sec	1/sec	1/sec	1/Mission
CRITERIA	EMP	EMP	EMP	OPER
TRANSFORMATION	REACTION TIME, FREQ, ACC, ME, PROPORTION	TIME, ME, MIN, MAX	TIME, ME, MIN, MAX	PROPORTION
CURRENTLY AVAILABLE		X	X	X
ACCESSIBLE	X			

TABLE XI: Contd

	MEASURE			
	Time-On-TP Error (Based on Planned Time of TP Passage Minus Actual Time of TP Passage)	Time-On-TP Computational Accuracy (Verbal Instructions to Pilot as to "FAST" or "SLOW")	Airspeed at TP Passage	Airspeed Between TPs
METHOD OF MEASUREMENT	E	O	E	E
MEASURE SEGMENT	TP	Entire leg	TP	Entire leg
SCALE	$\pm$500 sec	0/1	0 - 600 kts	0 - 600 kts
SAMPLING RATE	1/sec	1/sec	1/sec	1/sec
CRITERIA	EMP	OPER, EMP	EMP	EMP
TRANSFORMATION	ACC, MIN, ME, MO, MAX, RMS	PROPORTION	ACC, MIN, ME MAX	ACC, MIN, ME, MAX
CURRENTLY AVAILABLE		X	X	X
ACCESSIBLE	X			

TABLE XI: Contd

	MEASURE			
	Radar Scope Quality	Total Time Required to Identify TP with Radar and Cursor Intersection within Criterion Distance	Comparison Accuracy of Radar Return and Preplanned Chart and Quick & Dirty	Total Time Spent Slewing Cursors (Slew Control Stick and Radar Slew Button)
METHOD OF MEASUREMENT	O	E, O	O	E
MEASURE SEGMENT	Entire leg	Entire leg	Entire leg	Entire leg
SCALE	0/1	0 - 1000 sec	0/1	0 - 1000 sec
SAMPLING RATE	1/leg	1/sec	1/leg	1/sec
CRITERIA	OPER	EMP	OPER	EMP
TRANSFORMATION	PROPORTION	ACC, MIN, ME, MO, MAX	PROPORTION	TIME, MIN, ME, MO, MAX
CURRENTLY AVAILABLE	X	X	X	
ACCESSIBLE				X

TABLE XI: Contd

	MEASURE			
	Proportion Time Spent Slewing Cursors	Number of Times Cursors are Slewed	Number of Times Present Position is Updated Using Radar (CORRECT POS Button Activation)	Reaction Time to Navigational Equipment Failures (from Onset of Error; (Computer Error Light, INS Panel, etc.)
METHOD OF MEASUREMENT	E	E	E	E, O
MEASURE SEGMENT	Entire leg	Entire leg	Entire leg	Entire leg
SCALE	0 - 1	0 - 1000	0 - 100	0 - 1000 sec, 0/1
SAMPLING RATE	1/leg	1/sec	1/sec	1/sec
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	MIN, MO, MAX	MIN, ME, MO, MAX	MIN, ME, MO, MAX	TIME, MIN, ME, MAX
CURRENTLY AVAILABLE			X	(Computer Error Light)
ACCESSIBLE	X	X		X

TABLE XI: Contd

	MEASURE			
	Reaction Time to System Velocities out of Criterion Limits (from Onset to Correction Initiation)	Tracking Error (Between Cursor Intersection and Actual TP)	Tracking Error	Groundspeed
METHOD OF MEASUREMENT	E, O	E	E	E
MEASURE SEGMENT	Entire leg	Entire leg	TP	Entire leg
SCALE	0 - 1000 sec, 0/1	0 - 300K ft	0 - 300K ft	0 - 600 Kts
SAMPLING RATE	1/sec	1/sec	1/sec	1/sec
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	TIME, MIN, ME, MAX	ME, RMS	ME, RMS	ACC, MIN, ME, MAX
CURRENTLY AVAILABLE	X			
ACCESSIBLE		X	X	X

TABLE XI: Contd

	MEASURE			
	MA-1 Compass Needle Deflection Error	Magnetic Variation Setting Error (MAG VAR Display)	Altitude at TP Passage	Altitude Between TPs
METHOD OF MEASUREMENT	E	E	E	E
MEASURE SEGMENT	Entire leg	Entire leg	TP	Entire leg
SCALE	0 - 10 (units)	0 - 100 deg	0 - 50K ft	0 - 60K ft
SAMPLING RATE	1/sec	1/sec	1/sec	1/sec
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	ACC, ME, RMS	ACC, ME, RMS	ACC, MIN, ME, MAX	ACC, MIN, ME, MAX
CURRENTLY AVAILABLE			X	X
ACCESSIBLE	X	X		

TABLE XI: Contd

	MEASURE			
	Total Time on TP (Cursors within Criterion Limit of TP)	Total Time on TP	Number of Correct TP Identifications	Total Time Spent Performing Automatic/Manual Velocity Correct (VELOCITY CORRECT SW)
METHOD OF MEASUREMENT	E	E	O	E
MEASURE SEGMENT	Entire leg	TP	Mission	Entire leg
SCALE	0 - 1000 sec	0 - 1000 sec	0 - 15	0 - 1000 sec
SAMPLING RATE	1/sec	1/sec	1/leg	1/sec
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	MIN, ME, MO, MAX	MIN, ME, MO, MAX	MIN, ME, MAX, PROPORTION	MIN, MAX, ME
CURRENTLY AVAILABLE			X	X
ACCESSIBLE	X	X		

TABLE XI: Contd

	MEASURE			
	Proportion of Time Spent Performing AVC/MVC	Fuel Management (Estimated Minus Actual Fuel for Each Leg)	TP Accuracy (Actual distance Between Aircraft and TP)	Number of TPs Found (Within Criterion Distance for TP Accuracy)
METHOD OF MEASUREMENT	E	E	E	E
MEASURE SEGMENT	Entire leg	TP	TP	Mission
SCALE	0 - 1	$\pm$10K lbs	0 - 300K ft	0 - 15
SAMPLING RATE	1/leg	1/leg	1/TP	1/leg
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	MIN, MO, MAX	MIN, ME, MO, MAX PROPORTION	MIN, ME, MAX, RMS	MIN, ME, MAX PROPORTION
CURRENTLY AVAILABLE	X		X	X
ACCESSIBLE		X		

TABLE XI: Contd

	MEASURE			
	Proportion of TPs Found (within Criterion Distance for TP Accuracy)	Time Outside Criterion Flight Path	Number of Times Outside Criterion Flight Path	Ground Track Error
METHOD OF MEASUREMENT	E	E	E	E
MEASURE SEGMENT	Mission	Entire leg	Entire leg	Entire leg
SCALE	0 - 1	0 - 1000 sec	0 - 500	±100 deg
SAMPLING RATE	1/leg	1/sec	1/sec	1/sec
CRITERIA	EMP	EMP	EMP	EMP
TRANSFORMATION	MIN, ME, MAX	TIME, MIN, ME, MAX, PROPORTION	FREQ, MIN, ME MAX	ACC, ME, RMS
CURRENTLY AVAILABLE	X			X
ACCESSIBLE		X	X	

TABLE XI: Contd

	MEASURE			
	Criterion Flight Path Accuracy (Distance from Actual Position to Nearest Flight Path Boundary)			
METHOD OF MEASUREMENT	E			
MEASURE SEGMENT	Entire leg			
SCALE	0 - 300K ft			
SAMPLING RATE	1/sec			
CRITERIA	EMP			
TRANSFORMATION	MIN, ME, MO, MAX, RMS			
CURRENTLY AVAILABLE				
ACCESSIBLE	X			

3. Scale

Scale shows the numerical limits and units of the measure, e.g., seconds, feet, miles, or if units are unassigned, the values that the measure is assumed to take on over the measure segment. "0/1" defines a dichotomous "not occurred/occurred" situation. Subjective scales are listed separately (e.g., "effectiveness of communication").

4. Sampling Rate

The sampling rate is the rate of measurement defined by time or by measure segment. Determination was based on research by Vreuls and Cotton [1980].

5. Criteria

The recommended method of establishing performance criteria for the performance measure is defined as "EMP" for empirical, "OPER" for operational (subjective determination by fleet aircrew), or both.

6. Transformation

A recommended mathematical or statistical process for the measure is provided based on the literature review by Mixon and Moroney [1980]. A caution is provided that the determination of the measure's distribution would be in order before applying any transformations, as previously discussed in Chapter IV. Transformations are listed as TIME, FREQ (frequency), ACC (accuracy), ME (mean), MO (mode), PROPORTION (of successes to total number), MIN (minimum), MAX (maximum), RMS (root mean squared error), or N/A (not applicable; if measured for computational purposes only). Other transformations may be possible; see Table IV.

7. Currently Available

The measure exists within the A-6E WST PMS, as listed in Table VIII or Figure 5. If blank, the measure is not available.

8. Accessible

If not currently available in the A-6E WST PMS, a determination was made as to the feasibility of incorporating the measure into the existing WST PMS with a minor software change. If blank, major changes may be required in the WST PMS to facilitate the accessibility of the measure.

It is recognized that the resultant candidate measure set contains redundant and perhaps overlapping measures but analytical derivation is necessary before empirical analysis is possible. Only "objective" aspects of performance were listed; the determination of subjective components (i.e., motivation) and their measurement is a subject for further research.

B. ESTABLISHMENT OF PERFORMANCE STANDARDS

1. Radar Navigation Maneuver

The radar navigation maneuver consists of multiple "legs" of varying distances between predesignated turn points or targets that are selected using radar significant criteria. Each turn point (TP) must be reached within a criterion distance at a predetermined time, airspeed, heading, and altitude for the success of the air interdiction mission. Some TPs are more difficult than others to detect, identify and successfully fly over using the A-6E CAINS navigation system. Operationally, a TP is "crossed" or "reached" when the actual position of the TP passes more than ninety degrees abeam the actual aircraft position. This operational definition is independent of the DVRI moving bug cue used by the B/N, as based on cursor intersection on the perceived TP radar return. This operational definition of TP "passage" can easily be converted mathematically using Boolean functions for use in the A-6E WST performance measurement system.

Appropriate radar navigation routes can be planned by either FRS or Medium Attack Wing (MATWING) personnel and programmed into the WST. This task is simplified by the current capability of the simulator to facilitate preplanned mission routes for training purposes.

2. Simulator "Intruder Derby"

A competitive exercise is currently conducted on an annual basis using actual A-6E TRAM or CAINS aircraft that perform radar navigation maneuvers in both East and West Coast A-6 communities. A similar competition could be applied in the A-6E WST using the programmed radar navigation routes as previously discussed. Each route is then flown by fleet-experienced aircrew on a competitive basis under the cognizance of the appropriate MATWING command with performance measured by the proposed model. Using fleet aircrew that are carefully selected by individual squadron Commanding Officers almost guarantees motivated and skilled performance during the radar navigation routes due to the intrinsic importance placed upon the competition results as an aid in determining that A-6 squadron which is the most "excellent" in each MATWING community.

The performance results of the simulator "Intruder Derby" would be most useful for establishing standards of performance for each radar navigation route. Establishing standards in this manner for comparisons of performance to other groups is both feasible and operationally acceptable. The use of "ideal" flight path performance criteria lacks this acceptance criteria among operational FRS and fleet aviators, as "ideal" performance may not be achieved by even the most highly skilled aviator in a consistent manner. The fleet-established standards of performance would be carefully analyzed and performance limits set by operational personnel for those performance dimensions which are within or approaching the performance standards of the fleet.

C. MEASURE SELECTION TECHNIQUES

Various empirical methods and models have been formulated and applied toward reducing a list of candidate measures to a small, efficient set with the characteristics of reliability, validity, objectivity and timeliness. One technique, employed successfully by the Air Force for air combat maneuvering (ACM) performance measurement, used univariate and multivariate analysis techniques to find the smallest comprehensive set of measures which discriminated skill differences in "novice" and "expert" ACM pilots in one-versus-one free engagements [Kelly, et al., 1979]. Using multivariate analysis, correlational analysis, regression analysis, and ridge adjusted discriminant analysis, an original set of twenty-seven candidate performance measures were reduced to a final set of sixteen measures that were:

(1) Sensitive to differences in pilot ACM skill level.

(2) Diagnostic of performance proficiencies and deficiencies.

(3) Usable by instructor pilots and compatible with their judgements.

(4) Capable of providing results immediately after the end of the engagement.

(5) Compatible with current projected training and measurement hardware.

These statistical analysis techniques appear to be appropriate for application to the measurement of B/N performance during radar navigation in the A-6E WST. Computer programs have been developed and are available at minimum cost for possible software alterations to the current performance system in the WST.

D. EVALUATION METHODS

Once a small and efficient set of performance measures has been derived using suitable statistical techniques, measurement and evaluation of student performance may then occur.

The proposed performance measurement model uses the results of a student's performance compared to fleet performance as usable information for several decision levels. The task of operating the radar by the student on each leg of the radar navigation maneuver is measured and evaluated in terms of the underlying skill level of the student, leading to a decision of "proficient" or "not proficient" for that task. Decisions on quality of performance must also be made for global indices of navigational skill, e.g., fuel management or time-on-TP management. Students just learning the task are expected to be "not proficient" when compared to fleet performance on the same mission whereas students near the end of scheduled training are expected to meet fleet standards.

An evaluation technique proposed by Rankin and McDaniel [1980] that uses a sequential method of making statistical decisions has the capability of utilizing both objective and subjective performance results for more accurate and potentially less costly training evaluation. This decision model focuses on proportions of "proficient" trials, where "proficient" is determined by the instructor using either subjective evaluation or objective standards established prior to performance. The model sequentially samples performance during the training of a particular task or maneuver and uses the historically sampled performance results to eventually terminate training for that particular task or maneuver.

Figure 7 illustrates the proposed sequential sampling decision model using the task of navigating the A-6E aircraft to a radar navigation TP as an example. For this particular radar navigation route, eight TPs must each be navigated to within an empirically established criterion radial distance by the student. The instructor must use the performance measurement results to evaluate the student's actual performance on each TP and assign a "proficient" (P) or "not proficient" (1) score for each TP, where each TP is considered a trial. The figure shows "proficient" (P) trials plotted against total trials and indicates in this example that three TPs were successively and accurately navigated, followed by a missed fourth TP and ending with the remaining four TPs successfully navigated. The regions of "proficient," "undetermined," and "not proficient" are derived statistically; more detail on their actual calculation is presented in Appendix E. As can be shown in the figure, the student has "mastered" the important task of navigating to a TP on his eighth trial. This information can then be used by the training manager in deciding whether the student has actually mastered the task and needs to progress to more difficult tasks or the student has not mastered the task in a previously established, statistically-based number of training trials and needs remedial training for that task. For this example, the training manager could safely determine that the student is ready for training on tasks other than accurately navigating to a TP.

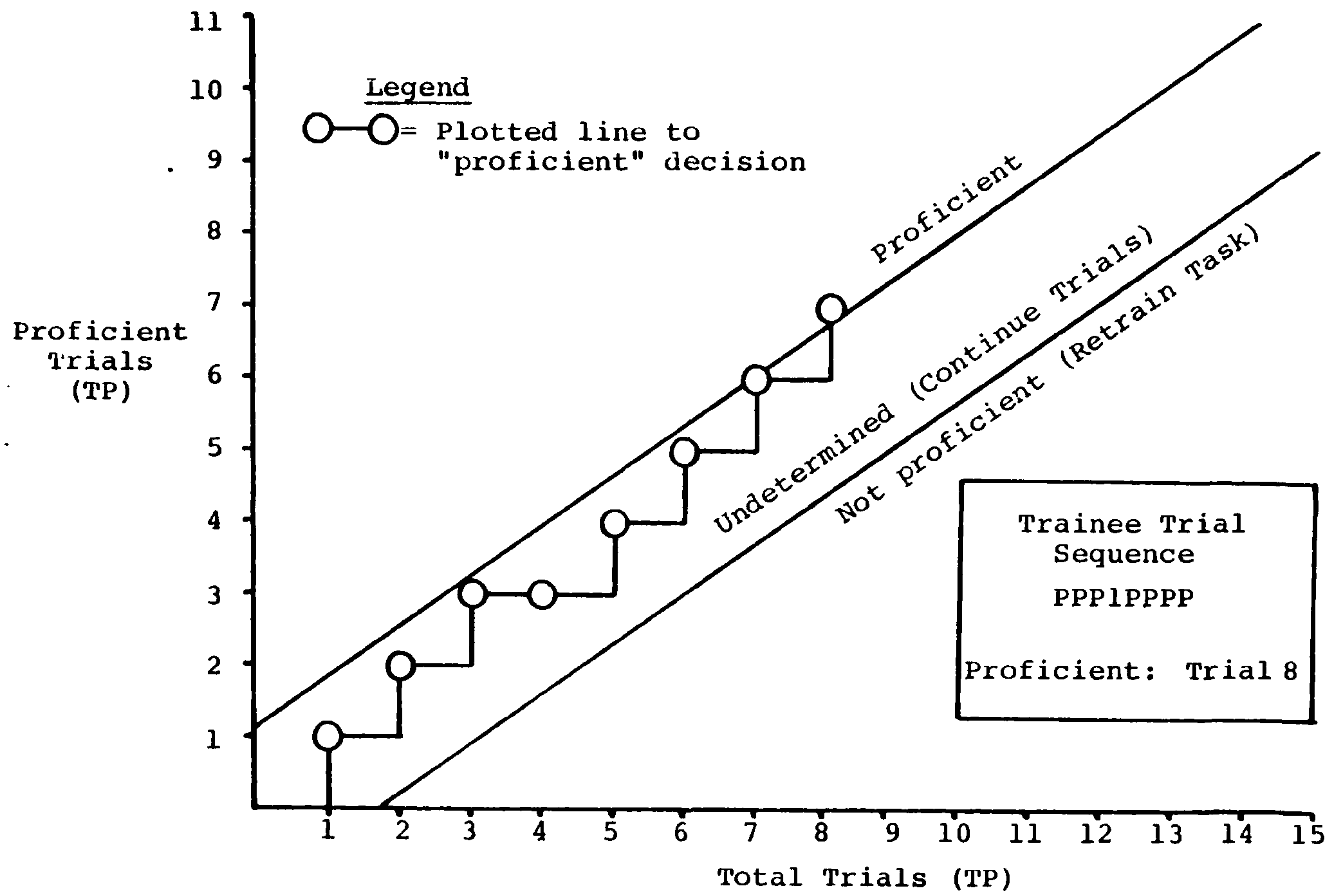


Figure 7. Sequential Sampling Decision Model for TP Navigation Task.

This sequential sampling decision model has been previously used in educational and training settings. Ferguson [1969] used the sequential test to determine whether individual students should be advanced or given remedial assistance after they completed instructional learning modules, and Kalisch [1980] employed the model for an Air Force Weapons Mechanics Training Course (63ABR46320) conducted at Lowry Air Force Base, Colorado. Both applications resulted in greater test efficiency than for tests composed of a fixed number of items and substantially reduced testing time.

As discussed in Chapter IV (Table V), the resulting costs associated with training manager decisional errors predicates that statistical and systematic methods be employed to measure and evaluate student performance. The sequential sampling plan accomplishes this by <u>fixing</u> the error rates (Types I and II as previously discussed in Table V) and allowing the number of trials to vary according to the performance demonstrated by the student. This evaluation decision model is currently being integrated into the training program of a helicopter FRS that uses only subjective determinations of "proficient" instructors [Rankin and McDaniel, 1980].

E. INFORMATION DISPLAYS

This section discusses some displays proposed for use by the instructor at the A-6E WST console for the purpose of performance evaluation. Several classes of performance measures are represented. Figure 8 shows a time activity record for the B/N's interactive control of the A-6E CAINS navigation system during one leg of a radar navigation maneuver. This type of display tells what, when, and for how long a particular equipment was being operated. From this, one may infer what particular task was being accomplished at a particular time during a radar navigation leg. Time activity records of fleet performance may be used as a performance standard by simply preparing a transparent overlay to show means and ranges of activity by the fleet performing the same radar navigation leg. This comparison provides diagnostic information for the individual student in regards to efficient or appropriate operation of the complex navigation system of the A-6E.

Specific tasks may be measured directly and displayed as shown in Figures 9 and 10. The tasks of time and fuel management are measured in the simulator by comparing planned time and fuel values with actual time and fuel values at each turn point, based on the premise of facilitating the input of the student-planned values by the instructor into the simulator computer prior to the simulated mission. Trends may show decisional errors not otherwise detectable by an instructor. Again, fleet performance standards can be compared by using simple overlays to an individual student's performance, as well as the performance of other students, to facilitate evaluation.

AD-A098 776 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 5/9

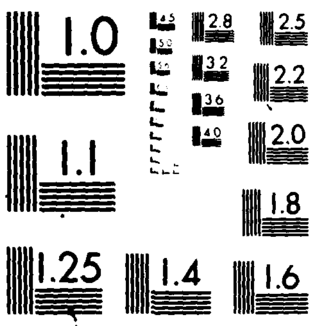
A MODEL TO MEASURE BOMBARDIER/NAVIGATOR PERFORMANCE DURING RADA--ETC(U)

MAR 81 T R MIXON

UNCLASSIFIED NL

3 of 4

AD A098 776



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

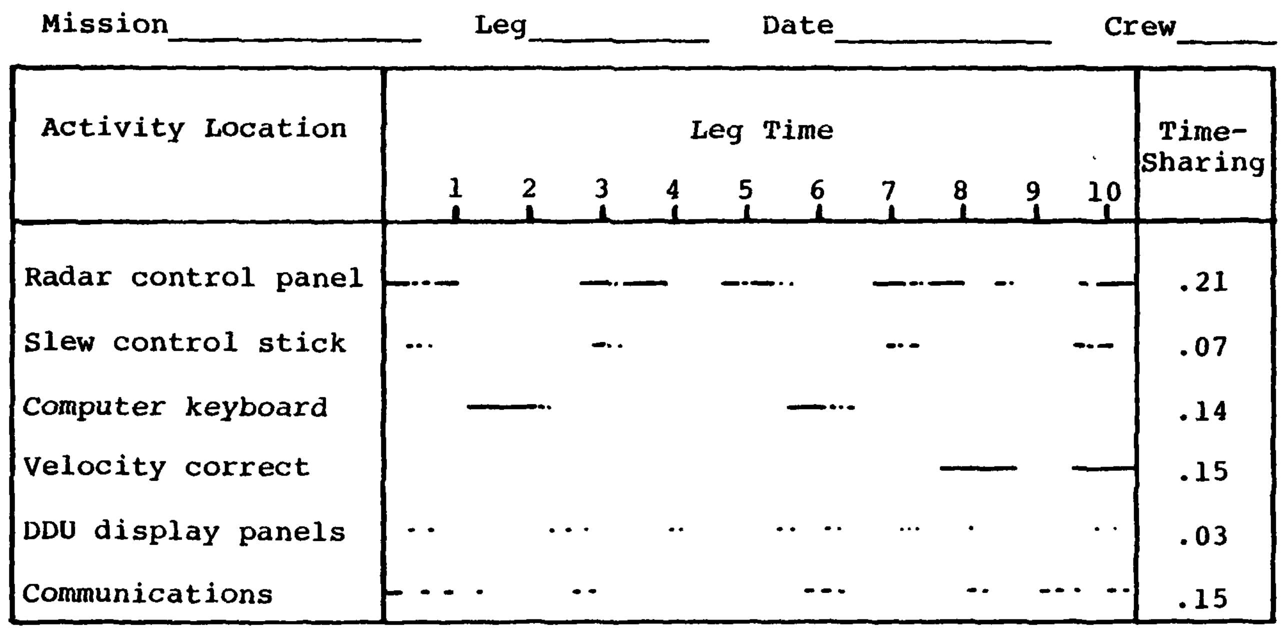


Figure 8. Time Activity Record Display.

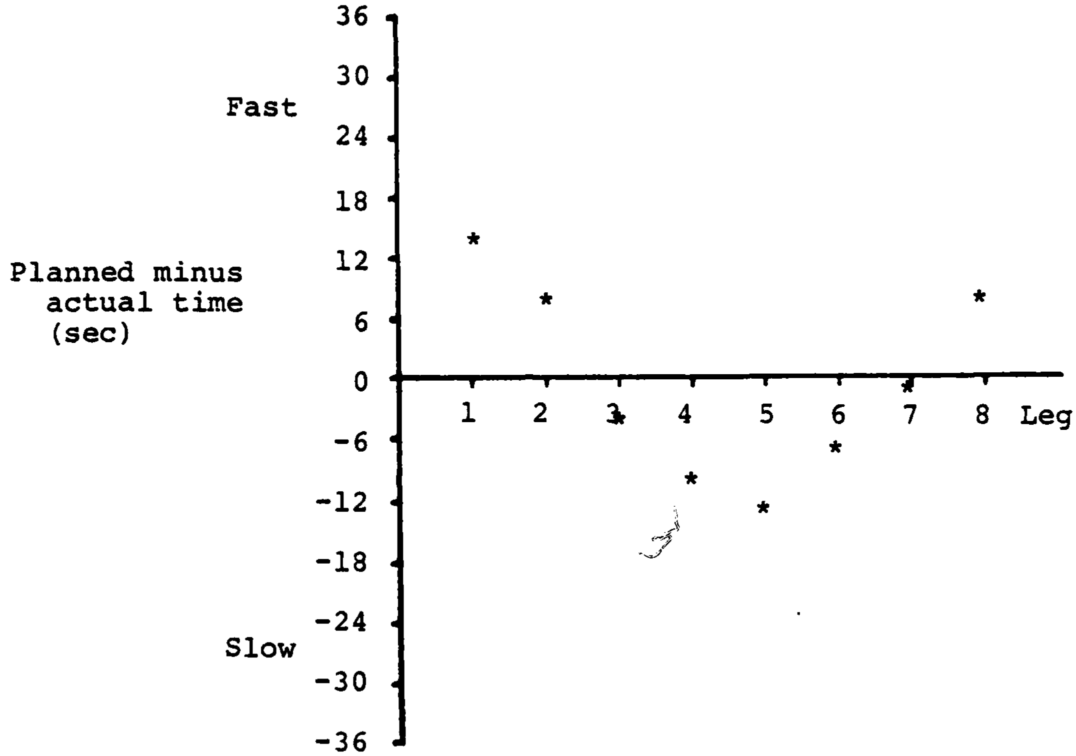


Figure 9. Time-on-TP Management Display.

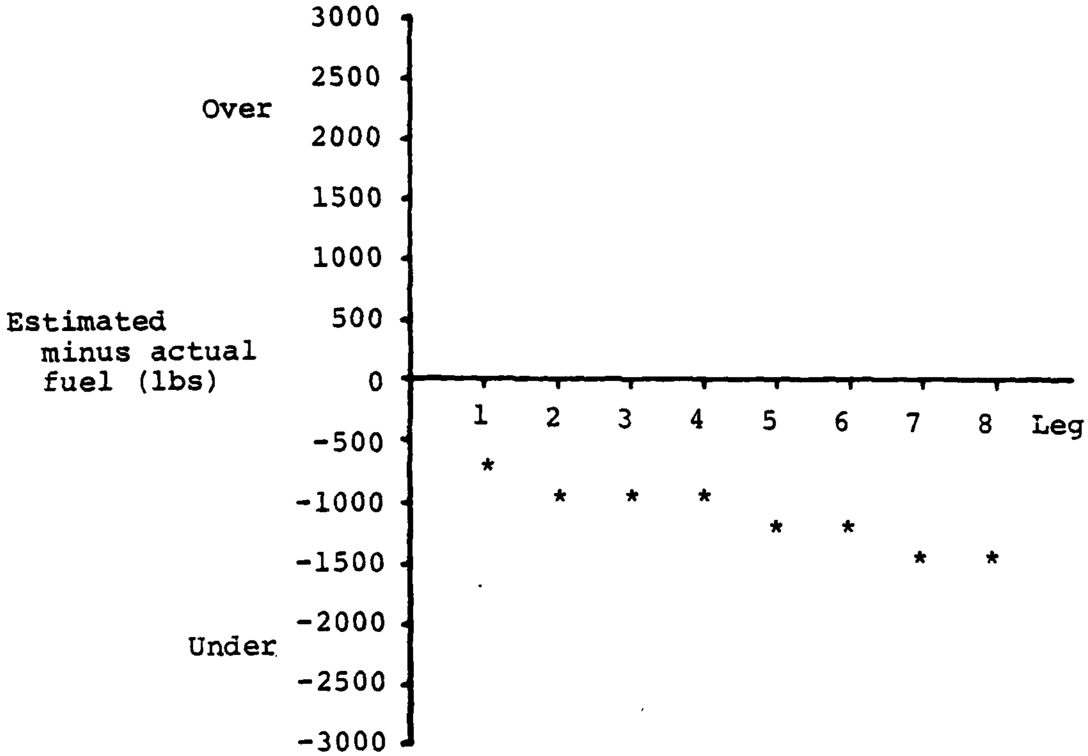


Figure 10. Fuel Management Display.

Figure 11 shows the usefulness of displaying the results of an overall measure of performance: radar navigation accuracy. The solid boundary lines between and surrounding each TP are performance standards established by fleet A-6E CAINS aircrew; in this situation a 90 percent confidence interval has been constructed about the mean flight path. Student navigational accuracy over the entire mission may be evaluated from this display as well as diagnostic navigational information for each leg or TP. This figure illustrates the performance of a student who has met fleet-established criterion limits for all leg and TP navigation portions of the route except TP number two.

Summarized performance of the entire mission is depicted by Table XII. Each turn point (TP) is evaluated for B/N equipment time-sharing activity on the previous leg, time-on-TP management, fuel management, heading accuracy (as related to planned run-in heading), navigational accuracy (reported as a "P" if within criterion limits or reported in miles from TP if not within limits), minimum altitude for the previous leg, and indicated airspeed (IAS) at the TP.

This brief discussion on information displays for B/N performance is not exhaustive and does not reflect what may be the most efficient and reliable measures for determining skill acquisition. Only statistical methods will produce those measures that should be displayed and used. The examples presented here are for illustrative purposes only.

Mission______________________ Date____________________ Crew____________________

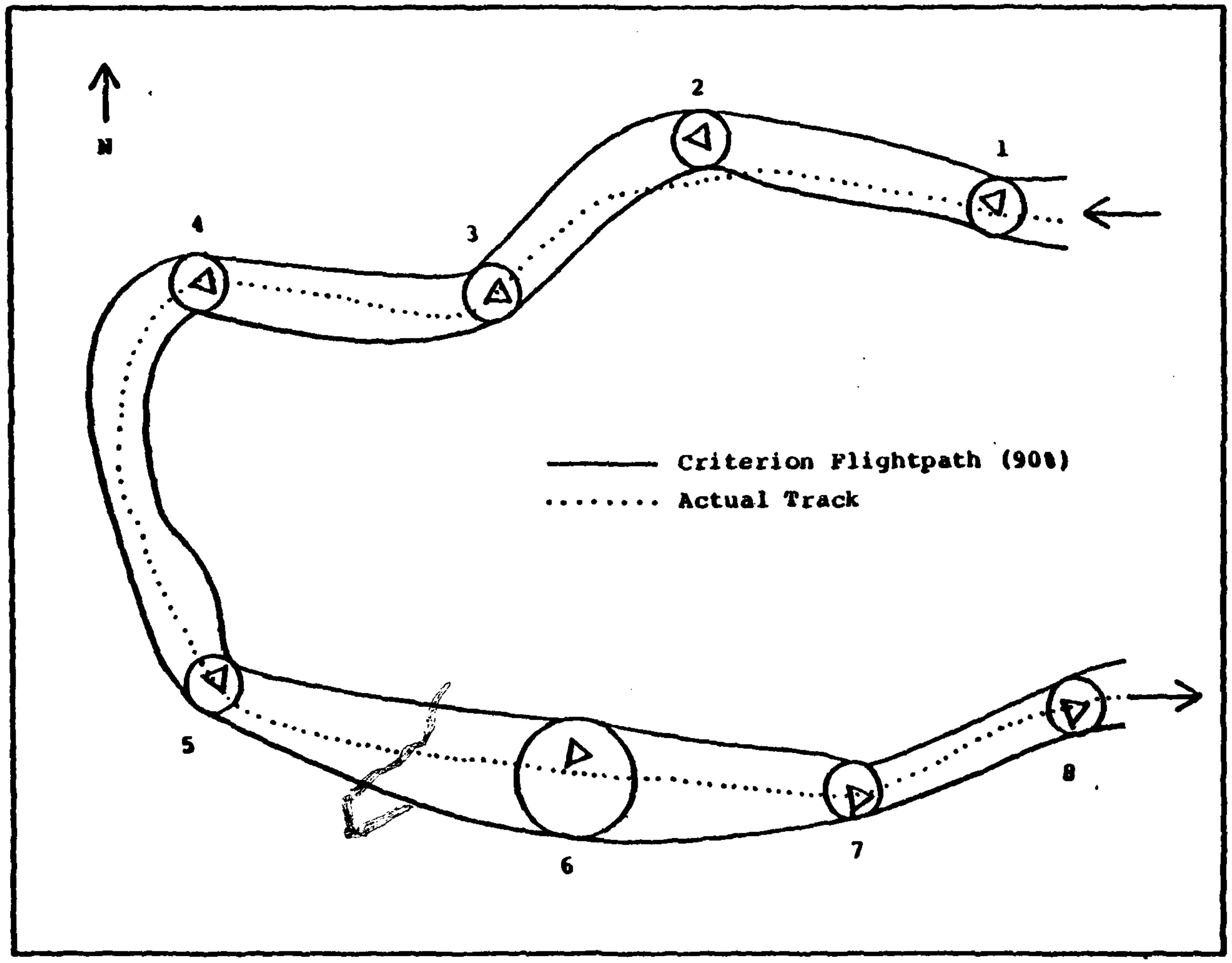


Figure 11. Radar Navigation Accuracy Display.

TABLE XII: RADAR NAVIGATION PERFORMANCE SUMMARY

TP	TP Name	Time-sharing Activity* 1	2	3	4	5	6	Time on TP	Fuel Mgt	Hdg	Nav	Alt	IAS
1	Ric Pol	.21	.07	.14	.15	.03	.15	11	-8	2	P	210	355
2	Catawba TPP	.27	.09	.13	.12	.02	.12	5	-12	1	1.1	250	355
3	Victoria HS	.22	.06	.15	.11	.01	.17	-6	-12	3	P	170	360
4	Hamlet RR POL	.34	.12	.12	.10	.04	.15	-12	-12	0	P	190	365
5	Danville PP	.21	.10	.01	.00	.01	.10	-6	-13		P	220	360
6	Smithfield	.17	.06	.00	.13	.02	.12	0	-13	-2	P	250	360
7	SBV HS	.28	.13	.00	.00	.02	.07	2	-15	0	P	260	360
8	Fay Term	.15	.05	.12	.00	.03	.03	8	-15	1	P	90	355
	μ	.23	.09	.08	.08	.02	.11	.25	-12	.25	.88	205	359
	σ	.06	.03	.07	.06	.01	.05	7.8	2.2	2		56	3.5

* 1 Radar control panel
2 Slew control stick
3 Computer keyboard
4 Velocity correct
5 DDU display panels
6 Communications

F. IMPLEMENTING THE MODEL

A model for measuring B/N performance during radar navigation in the A-6E WST has been designed and proposed for use by the East and West Coast A-6 Fleet Replacement Squadrons. The final model design was predicated on: (1) implementing the model at minimum cost, (2) utilizing existing computer algorithms and software that have been validated, and (3) requiring no additional personnel to operate the model after implementation. Some software changes are necessary, but they appear to be minor in light of the 2F114 design specifications for currently accessible programs. Some translation may be necessary due to different computer languages but these are feasible alternatives given the implications for reducing training costs and increasing the effectiveness of both the instructor and the simulator. Additionally, a computer-managed system will be necessary for implementing the sequential sampling decision model. Currently available desk-top computers could accomplish this function assuming the simulator's computer capacity was fully utilized after the measurement portion of the model was installed.

Objective performance measurement provides useful information necessary for training evaluation and control. Performance measurement models that incorporate this powerful technique can increase simulator and instructor effectiveness, reduce training costs, and may contribute toward reducing accidents attributed to "unskilled" aviators.

Implementation of these systems appears to be cost-effective in view of the potential savings from their effective utilization.

VIII. SUMMARY

The purpose of this thesis was to model a performance measurement system for the Bombardier/Navigator (B/N) Fleet Replacement Squadron (FRS) student during the radar navigation maneuver in the A-6E Weapon System Trainer (WST, device 2F114) that would best determine student skill acquisition and would incorporate the advantages of both objective and subjective aircrew performance measurement methods. This chapter is provided as a compendium due to the extensive material covered and assumes reader unfamiliarity of previous chapters.

A. STATEMENT OF PROBLEM

Traditional and current FRS student performance measurement and assessment in the A-6E WST by an instructor is mostly subjective in nature with disadvantages of low reliability, lack of established performance standards, and human perceptual measurement inadequacies. The recently delivered A-6E WST has the capability to objectively measure student performance but is not being utilized in this fashion due to the lack of an operational performance measurement system that incorporates the characteristics of objective performance measurement and still retains the valuable judgement and experience of the instructor as a measuring and evaluating system component. Objectivity in performance measurement is a highly desirable component of the performance measurement and evaluation process that enables the establishment of performance standards, increases instructor and simulator effectiveness, and fulfills the requirements of Department of Defense policy.

B. APPROACH TO PROBLEM

Designing a model to measure B/N performance during radar navigation in the A-6E WST necessarily assumed: (1) the A-6E WST realistically duplicated the A-6E aircraft in both engineering and mission aspects, (2) little variability in overall A-6E crew-system performance is attributable to the pilot, (3) that results from pilot performance measurement literature were applicable to the B/N, (4) that a mathematical relationship existed between some aspects of B/N behavior and performance measurement and evaluation, and (5) competitively selected, motivated and experienced A-6E fleet aircrew exhibit advanced skill or "proficiency" characterized by minimum effort and consistent responses ordinarily found in actual aircraft flight. The methodology used in formulating a model to measure B/N performance was based on an extensive literature review of aircrew performance measurement from 1962-1980 and an analytical task analysis of the B/N's duties. After selection of the Air Interdiction scenario and turn point-to-turn point radar navigation flight segment, the review concentrated on aircrew performance measurement research which emphasized navigation, training, and skill acquisition. A brief review was presented of the concepts of performance measurement and evaluation including measure types, reliability, validity, measure selection criteria, performance standards, aviation measures of effectiveness and types of evaluation within the framework of aircrew training. A model was then formulated to illustrate the relationship among student B/N skill acquisition, the radar navigation task, and performance measurement and evaluation. Candidate measures for navigation training and the radar navigation flight segment were identified from an original listing of 182 performance measures from previous aircrew performance measurement research. The task analysis was performed to identify skills and knowledge required of the B/N and to identify candidate performance measures for both B/N skill acquisition and the radar navigation segment. A Mission Time Line Analysis (MTLA) was conducted to identify B/N tasks critical to performance. A model was then formulated to illustrate A-6E crew-system interaction and the complexity involved in measuring B/N performance. Generic aircrew performance measurement system concepts were reviewed for the training environment. Current performance measurement and evaluation practices of the A-6 FRS for the B/N student in the WST were reviewed as well as the current objective performance capabilities of the WST. A final list of candidate measures was presented that had met selection criteria of face validity, ease of use, instructor and student acceptance, and appropriateness to training.

C. FORMULATION OF THE MODEL

The purpose of measuring B/N performance during radar navigation in the WST was to provide objective, reliable, and valid information for accurate decision-making about B/N skill acquisition to the FRS instructor and training manager. The model developed candidate measures (Table XI) that determined what to measure, observation method, when to measure, scaling, sampling frequency, criteria establishment, applicable transformations, current availability in the A-6E WST, and accessibility of the measure if not currently available. After operationally defining the radar navigation segment, a proposal was made to conduct an annual competitive exercise in the A-6E WST under the cognizance of the appropriate Medium Attack Wing (MATWING) command utilizing A-6E fleet squadron aircrew. A-6E fleet squadron personnel would fly preprogrammed radar navigation routes while their performance was measured using the candidate measures previously developed. The results from the proposed competitive exercise were cited as being useful and operationally acceptable for establishing standards of performance for each radar navigation route since the selected aircrew would be highly motivated and fleet-experienced. Statistical techniques for reducing the initial candidate measures for B/N radar navigation performance were reviewed and evaluated with respect to skill acquisition, and a multivariate discriminant analysis model was selected as applicable due to the model's utility and previous practical development and applications.

The final part of the performance measurement model proposed an evaluation application which used dichotomous results of student performance compared to fleet performance as information for several decision levels. Performance results of the student for the majority of the statistically reduced performance measures can be dichotomized by the instructor as either "proficient" (skilled) or "not proficient" (unskilled) based on the empirically established objective performance standards or operationally-defined subjective performance standards. An evaluation model developed by Rankin and McDaniel [1980] that used a sequential method of making statistical decisions incorporating the dichotomized results of student performance was adapted and modified for determining successful completion of task training for the B/N based on both objective and subjective performance measurement and evaluation. The sequential sampling model was selected due to its inherent power to fix decisional error rates, previous practical developments, and potential for reducing training costs. Several informational displays specific to the B/N radar navigation segment and based on hypothetical model results were presented. Model implementation was discussed with regards to personnel, costs, A-6E WST software changes, and effective training control.

D. IMPACT TO THE FLEET

The performance measurement model as outlined in this thesis is specific to measuring A-6E B/N performance during radar navigation but has generic qualities applicable to aircrew members of any aircraft. The advantages of objective measurement in the form of reduced paperwork, permanent performance records, established performance standards, diagnostic and timely information, and high reliability are fulfilled. At the same time, subjective measurements for those aircrew behavioral aspects that currently defy objective measurement are made using the experienced simulator mission instructor, who also remains as the final decision-maker on whether or not a student has demonstrated task performance that reflects an acquired skill for that task.

The application of the model to individual aircrew readiness, fleet squadron unit readiness, and selection of individual aircrew teams for multi-crew aircraft appears to be feasible and operationally acceptable. The model has potential utility in tactics development, accident prevention, predictive performance, and proficiency training of reserve aviators. Almost certainly some reduction in aircrew training costs and an increase in instructor, simulator, and training program effectiveness would be realized.

BIBLIOGRAPHY

Alluisi, E.A., "Methodology in the Use of Synthetic Tasks to Assess Complex Performance," Human Factors, v. 9(4): p. 375-384, August 1967.

Anderson, R.C. and Faust, G.W., Educational Psychology: The Science of Instruction and Learning, Dodd, Mead & Company, 1974.

Angell, D., Shearer, J.W., and Berliner, D.C., Study of Training Performance Evaluation Techniques, U.S. Naval Training Device Center, Port Washington, NY, NAVTRADEVCEN 1449-1, October 1964, AD 609605.

Bergman, B., and Siegel, A., Training Evaluation and Student Achievement Measurement: A Review of the Literature, Air Force Human Resources Laboratory, Lowry AFB, CO 80230, AFHRL-TR-3, January 1972, AD 747040.

Beverly, R.F., A Measure of Radar Scope Interpretation Ability RSI Test #1, Human Resources Research Center, Lackland AFB, San Antonio, TX, Research Note AO: 52-9, December 1952, AD 7956.

Billings, C.E., "Studies of Pilot Performance: I. Theoretical Considerations," Aerospace Medicine, v. 39: p. 17-19, January 1968.

Billings, C.E., Eggspuehler, J.J., Gerke, R.J., and Chase, R.C., "Studies of Pilot Performance: II. Evaluation of Performance During Low Altitude Flight in Helicopters," Aerospace Medicine, v. 39: p. 19-31, January 1968.

Bilodeau, E.A., Acquisition of Skill, Academic Press, 1966.

Bilodeau, E.A. (Editor), Principles of Skill Acquisition, Academic Press, 1969.

Bittner, A.C. Jr., Statistical Tests for Differential Stability, Proceedings of the 23rd Annual Meeting of the Human Factors Society, p. 541-545, October 1979.

Blum, A., Melynis J., Flader, D., Beardsley, W., Ramberg, T., Criteria Report for A-6E Weapon System Trainer Device 2F114, Final, Vol. I, Grumman Aerospace Corporation, TR-T01-CD-A005-04, January 1977.

Blum, A., Lovrecich, J., Force, J., Cohen, L., Happ, N., Configuration Report for A-6E Weapon System Trainer Device 2F114, Final, Vol. I, Grumman Aerospace Corporation, TR-T01-CD-A001-06, May 1977.

Blum, A., Lovrecich, J., Force, J., Cohen, L., Happ, N., Configuration Report for A-6E Weapon System Trainer Device 2F114, Final, Vol. II, Grumman Aerospace Corporation, TR-T01-CD-A001-06, May 1977.

Bowen, H.M., Bishop, E.W., Promisel, D., and Robins, J., Study, Assessment of Pilot Proficiency, U.S. Naval Training Device Center, Port Washington, NY, NAVTRADEVCEN 1614-1, August 1966, AD 637659.

Brictson, C.A., Human Factors Research on Carrier Landing System Performance, Final Report (1966-1971), Office of Naval Research, July 1971, AD 733703.

Buckhout, R., A Bibliography on Aircrew Proficiency Measurement, Aerospace Medical Division, Wright-Patterson AFB, OH, MRL-TDR-62-49, May 1962, AD 283545.

Buckhout, R., and Cotterman, T.E., Considerations in the Design of Automatic Proficiency Measurement Equipment in Simulators, 6570th Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH, AMRL Memorandum P-40, June 1963, AD 417425.

Campbell, S., Task Listing, A-6E (TRAM) Aircraft, Pilot Procedures, Vol. I, Grumman Aerospace Corporation, Report ISD (SAT) 0003-1, September 1975.

Campbell, S., and Sohl, D., Task Listing, A-6E (TRAM) Aircraft, Airframe Emergencies and System Malfunctions, Vol. III, Grumman Aerospace Corporation, Report ISD (SAT) 0003-3A, October 1975.

Campbell, S., Graham, G., and Hersh, S., Specific Behavioral Objectives and Criterion Test Statements, Pilot Procedures A-6E (TRAM) Aircraft, Vol. I, Grumman Aerospace Corporation, Report ISD (SAT) 0003-4, December 1975.

Campbell, S.C., Feddern, J., Graham, G., and Morganlander, M., Design of A-6E TRAM Training Program Using ISD Methodology. Phase I Final Report, Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN 75-C-0099-1, May 1976. Also titled as A-6E Systems Approach to Training. Phase I Final Report, February 1977.

Carter, V.E., Development of Automated Performance Measures for Introductory Air Combat Maneuvers, Proceedings of the 21st Annual Meeting of the Human Factors Society, p. 436-439, October 1977.

Chapanis, A., Theory and Methods for Analyzing Errors in Man-Machine Systems, from Sinaiko, H.W. (Editor), Selected Papers on Human Factors in the Design and Use of Control Systems, Dover Publications, Inc., p. 42-66, 1961.

Charles, J.P., Instructor Pilot's Role in Simulator Training Phase III), Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN 76-C-0034-2, June 1978, AD A061732.

Chiles, W., Objective Methods for Developing Indices of Pilot Workload, Federal Aviation Administration, Oklahoma City, OK 73125, FAA-AM-77-15, July 1977, AD A044556.

Christiansen, J.M., and Mills, R.G., "What Does the Operator Do in Complex Systems," Human Factors, v. 9(4): p. 329-340, August 1970.

Coburn, R., Human Engineering Guide to Ship System Development, Naval Electronics Laboratory Center, San Diego, CA 92152, NLEC/TD 278, October 1973.

Comptroller General of the United States, Greater Use of Flight Simulators in Military Pilot Training Can Lower Costs and Increase Pilot Proficiency, General Accounting Office, Washington, D.C. 20548, B-157905, 9 August 1973.

Connelly, E.M., Bourne, F.J., and Loental, D.G., Computer-Aided Techniques for Providing Operator Performance Measures, Air Force Human Resources Laboratory, Wright-Patterson AFB, OH 45433, AFHRL-TR-74-87, December 1974, AD A014330. See also Proceedings of the 18th Annual Meeting of the Human Factors Society, Huntsville, AL, p. 359-367, October 1974.

Connelly, E.M., Bourne, F.J., Loental, D.G., Migliaccio, J.S., Burchick, D.A., and Knoop, P.A., Candidate T-37 Pilot Performance Measures for Five Contact Maneuvers, Air Force Human Resources Laboratory, Wright-Patterson AFB, OH 45433, AFHRL-TR-74-88, December 1974, AD A014331.

Cotterman, T.E., and Wood, M.E., Retention of Simulated Lunar Landing Mission Skills: A Test of Pilot Reliability, Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH 45433, AMRL-TR-66-222, April 1967, AD 817232.

Cureton, E.E., Validity, from Lindquist, E.F. (Editor), Educational Measurement, George Banta Publishing Company, 1951.

Daniel, R.S., and Eason, R.G., Error in Aiming Point Identification on a Radar Bombing Mission, Air Force Personnel and Training Research Center, Mather AFB, CA, Project No. 7711, April 1954, AD 36050.

Danneskiold, R.D., Objective Scoring Procedure for Operational Flight Trainer Performance, U.S. Naval Special Devices Center, Port Washington, NY, SPECDEVCEN 999-2-4, February 1955, AD 110925.

Davis, R.H., and Behan, R.A., Evaluating System Performance in Simulated Environments, from Gagne, R.M. (Editor), Psychological Principles in System Development, Holt, Rinehart, & Winston, Inc., 1966.

Dawes, R.M., "The Robust Beauty of Improper Linear Models in Decision Making," American Psychologist, v. 34(7): p. 571-582, July 1979.

Deberg, O.H., A Scoring Algorithm for Assessing Air-to-Air Combat Performance. Phase I: A Factor Analysis, Air Command and Staff College, Maxwell AFB, AL 36112, Report No. 0590-77, May 1977, AD B019569.

Demaree, R.G., Norman, D.A., and Matheny, W.G., An Experimental Program for Relating Transfer of Training to Pilot Performance and Degree of Simulation, U.S. Naval Training Device Center, Port Washington, NY, NAVTRADEVCEN 1388-1, June 1965, AD 471806.

Department of Defense Military Specification MIL-H-46855B, Human Engineering Requirements for Military Systems, Equipment and Facilities, January 1979.

Dixon, W.J., and Massey, F.J. Jr., Introduction to Statistical Analysis, 3rd ed., McGraw-Hill, 1969.

Doll, R.E., Lane, N.E., Shannon, R.H., and Clisham, W.F. Jr., Naval Flight Officer Function Analysis Volume V: A-6 Bombardier/Navigator (B/N), Naval Aerospace Medical Research Laboratory, Pensacola, FL, December 1972.

Engler, H.F., Davenport, E.L., Green, J., and Sears, W.E. III, Human Operator Control Strategy Model, Air Force Human Resources Laboratory, Wright-Patterson AFB, OH 45433, AFHRL-TR-79-60, April 1980, AD A084695.

Ericksen, S.C., A Review of the Literature on Methods of Measuring Pilot Proficiency, Human Resources Research Center, Lackland AFB, San Antonio, TX, Research Bulletin 52-25, August 1952, ATI No. 169181.

Farrell, J.P., Measurement Criteria in the Assessment of Helicopter Pilot Performance, Proceedings of a Conference on Aircrew Performance in Army Aviation, held at U.S. Army Aviation Center, Fort Rucker, Alabama on November 27-29, 1973, Office of the Chief of Research, Development and Acquisition (Army) and U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA 22209, July 1974, p. 141-148, AD A001539.

Farrell, J.P., and Fineberg, M.L., "Specialized Training versus Experience in Helicopter Navigation at Extremely Low Altitudes," Human Factors, v. 18(3): p. 305-308, June 1976.

Federman, P.J., and Siegel, A.I., Communications as a Measurable Index of Team Behavior, U.S. Naval Training Device Center, Port Washington, NY, NAVTRADEVCEN 1537-1, October 1965, AD 623135.

Ferguson, R., The Development, Implementation, and Evaluation of A Computer-Assisted Branch Test for A Program of Individually Prescribed Instruction, Ph.D. Dissertation, University of Pittsburgh, 1969.

Fineberg, M.L., Navigation and Flight Proficiency under Nap of the Earth Conditions as a Function of Aviator Training and Experience, Proceedings of the 18th Annual Meeting of the Human Factors Society, Huntsville, AL, p. 249-254, October 1974.

Fineberg, M.L., Meister, D., and Farrell, J.P., An Assessment of the Navigation Performance of Army Aviators Under Nap-of-the-Earth Conditions, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA 22333, Research Report 1195, August 1978, AD A060563.

Fitts, P.M., Psychomotor Skills and Behavior Modification, Academic Press, 1965.

Forrest, F.G., Develop An Objective Flight Test for the Certification of A Private Pilot, Federal Aviation Administration, Washington, D.C. 20590, FAA-DS-70-17, May 1970, AD 708293.

Geiselhart, R., Kemmerling, P., Cronburg, J.G., and Thorburn, D.E., A Comparison of Pilot Performance Using a Center Stick VS Sidearm Control Configuration, Aeronautical Systems Division, Wright-Patterson AFB, OH, Technical Report ASD-TR-70-39, November 1970, AD 720846.

Geiselhart, R., Jarboe, J.K., and Kemmerling, P.T. Jr., Investigation of Pilots' Tracking Capability Using a Roll Command Display, Aeronautical Systems Division, Wright-Patterson AFB, OH, ASD-TR-71-46, December 1971, AD A009590.

Geiselhart, R., Schiffler, R.J., and Ivey, L.J., A Study of the Task Loading Using a Three Man Crew on a KC-135 Aircraft, Aeronautical Systems Division, Wright-Patterson AFB, OH 45433, ASD-TR-76-19, October 1976, AD A044257.

Gerlach, S., Cues, Feedback and Transfer in Undergraduate Pilot Training, Air Force Office of Scientific Research, Bolling AFB, DC 20332, AFOSR-TR-76-0056, August 1975, AD A033219.

Glauser, R., and Klaus, D.J., Proficiency Measurement: Assessing Human Performance, from Gagne, R.M. (Editor), Psychological Principles in System Development, Holt, Rinehart & Winston, Inc., 1966.

Gomer, F.E., Beideman, L.R., and Levine, S.H., The Application of Biocybernetic Techniques to Enhance Pilot Performance During Tactical Missions, Defense Advanced Research Projects Agency, MDC E2046, October 1979.

Graham, G., Hanish, J., and Hersh, S., Specific Behavioral Objectives and Criterion Test Statements, Airframe Emergencies and System Malfunctions, A-6E (TRAM) Aircraft, Vol. III, Grumman Aerospace Corporation, Report ISD (SAT) 0003-6, December 1975.

Greening, C.P., Performance Measurement and Prediction in Air-to-Ground Target Acquisition, Naval Weapons Center, China Lake, CA 93555, NWC TP 5747, April 1975.

Greer, G.D. Jr., Smith, D., Hatfield, J.L., Colgan, C.M., and Duffy, J.O., PPDR Handbook. Use of Pilot Performance Description Record in Flight Training Quality Control, Human Resources Research Office, Fort Rucker, AL, Department of the Army, Washington, DC 20310, December 1963, AD 675337.

Grodsky, M.A., "The Use of Full Scale Mission Simulation for the Assessment of Complex Operator Performance," Human Factors, v. 9(4): p. 341-348, August 1967.

Hanish, J., and Feddern, J., Specific Behavioral Objectives and Criterion Test Statements, Bombardier/Navigator Procedures, A-6E (TRAM) Aircraft, Vol. II, Grumman Aerospace Corporation, Report ISD (SAT) 0003-5, December 1975.

Haygood, R.C., Leshowitz, B., Parkinson, S.R., and Eddowes, E.E., Visual and Auditory Information Processing Aspects of the Acquisition of Flying Skill, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-74-79, December 1974, AD A007721.

Hitch, C., "Sub-optimization in Operations Problems," Journal of the Operations Research Society of America, v. 1(3): p. 87-99, May 1953.

Hitchcock, L. Jr., The Analysis of Human Performance within the Operational Flight Environment, Proceedings, Annual AGARD Meeting of the Aerospace Medical Panel, Toronto, Canada, Advisory Group for Aerospace Research and Development, Paris, France, AGARD Conference Proceedings No. 14, p. 9-14, September 1966, AD 661165.

Hulin, C.L., and Alvares, K.M., An Evaluation of Three Possible Explanations of the Temporal Decay in Predicting Pilot Proficiency, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL TR-71-5, February 1971, AD 731191

Hulin, C.L., and Alvares, K.M., Three Explanations of Temporal Changes in Ability-Skill Relationships: Literature Review and Theoretical Analysis, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL TR-71-6, February 1971, AD 732612.

Jensen, R.S., Vander Kolk, R.J., and Roscoe, S.N., Pilotage Error in Area Navigation: Effects of Symbolic Display Variables, Federal Aviation Administration, Washington, DC 20590, FAA-RD-72-29, January 1972, AD 750178.

Johnson, H.M., and Boots, M.L., Analysis of Ratings in the Preliminary Phase of the CAA Training Program, CAA Division of Research Report No. 21, 1943.

Jones, M.D., "Rate and Terminal Processes in Skill Acquisition," Journal of Psychology, v. 83: 1970.

Jones, R.A., The F-111D Simulator Can Reduce Cost and Improve Aircrew Evaluation, Air War College, Maxwell AFB, AL 36112, Professional Study No. 5959, April 1976, AD B010452.

Kalisch, S.J., Computerized Instructional Adaptive Testing Model: Formulation and Validation, Air Force Human Resources Laboratory, Brooks AFB, TX 78235, AFHRL-TR-79-33, February 1980.

Kelley, C.R., and Wargo, M.J., Adaptive Techniques for Synthetic Flight Training Systems, Naval Training Device Center, Orlando, FL, NAVTRADEVCEN 68-C-0136-1, October 1968, AD 678536.

Kelly, M.J., Wooldridge, L., Hennessy, R.T., Vreuls, D., Barnebey, S.F., Cotton, J.C., and Reed, J.C., Air Combat Maneuvering Performance Measurement, Naval Training Equipment Center, Orlando, FL 32813, and Air Force Human Resources Laboratory, Brooks AFB, TX 78235, NAVTRAEQUIPCEN IH-315/AFHRC-TR-79-3, September 1979; see also Proceedings of the 23rd Annual Meeting of the Human Factors Society, p. 324-328, 1979.

Kemmerling, P.T. Jr., The Impact of Full-Mission Simulation on the TAC Combat Crew Training System, Air War College, Maxwell AFB, AL 36112, Professional Study No. 5655, April 1975, AD B004457.

Klier, S., and Gage, H., Motion Factors in Flight Simulation, Naval Training Device Center, Orlando, FL 32813, NAVTRADEVCEN 68-C-0007-1, December 1970, AD 880341.

Knoop, P.A., Programming Techniques for the Automatic Monitoring of Human Performance, Aerospace Medical Research Laboratory, Wright-Patterson SFB, OH 45433, AMRL-TR-66-16, April 1966, AD 637454.

Knoop, P.A., Development and Evaluation of a Digital Computer Program for Automatic Human Performance Monitoring in Flight Simulator Training, Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH 45433, AMRL-TR-67-97, August 1968, AD 675542.

Knoop, P.A., and Welde, W.L., Automated Pilot Performance Assessment in the T-37: A Feasibility Study, Air Force Human Resources Laboratory, Advanced Systems Division, AFHRL/AS, Wright-Patterson AFB, OH 45433, AFHRL-TR-72-6, April 1973, AD 766446; see also Human Factors, v. 15(6): p. 583-597, December 1973.

Krendel, E.S., and Bloom, J.W., The Natural Pilot Model for Flight Proficiency Evaluation, U.S. Naval Training Device Center, Port Washington, NY, NAVTRADEVCEN 323-1, April 1963, AD 410805.

Leshowitz, B., Parkinson, S.R., and Waag, W.L., Visual and Auditory Information Processing in Flying Skill Acquisition, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-74-103, December 1974, AD A009636.

Lewis, R.E.F., Navigation of Helicopters in Slow and Very Low Flight. A Comparison of Solo and Dual Pilot Performance, Proceedings, Annual AGARD Meeting of the Aerospace Medical Panel, Toronto, Canada, Advisory Group for Aerospace Research and Development, Paris, France, AGARD Conference Proceedings No. 14, September 1966, p. 29-34, AD 661165.

Lindsay, G.F., Statistical Considerations in MX Site Characterization, Systems Exploration, Inc., Report No. B3440-79462-H53BAXSX37718-H-2590D, November 1979.

Locke, E.A., Zavala, A., and Fleishman, E.A., "Studies of Helicopter Pilot Performance: II. The Analysis of Task Dimensions," Human Factors, v. 7(3): p. 285-301, June 1965.

Lorge, I., The Fundamental Nature of Measurement, from Lindquist, E.F. (Editor), Educational Measurement, George Banta Publishing Company, 1951.

Marks, M.R., Development of Human Proficiency and Performance Measures for Weapon Systems Testing, Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH, ASD Technical Report 61-733, December 1961, AD 272975.

Matheny, W.G., Patterson, G.W. Jr., and Evans, G.I., Human Factors in Field Testing, Office of Naval Research, Washington, DC, LS-ASR-70-1, May 1970, AD 716438.

McCormick, E.J., Human Factors in Engineering and Design, McGraw-Hill, 4th Ed., 1976.

McCoy, W.K. Jr., "Problems of Validity of Measures Used in Investigating Man-Machine Systems," Human Factors, v. 5(4): p. 373-377, August 1963.

McDowell, E.D., The Development and Evaluation of Objective Frequency Domain Based Pilot Performance Measures in ASUPT, Air Force Office of Scientific Research, Bollings AFB, DC 20332, AFOSR-TR-78-1239, April 1978, AD A059477.

McMinn, V.L., "A Subjective Leap," Approach, v. 26(7): p. 24-25, January 1981.

Meister, D., and Rabideau, G.F., Human Factors Evaluation in System Development, Wiley and Sons, 1965.

Miller, N.E., Galt, W.E., and Gershenson, C.P., A Large Scale Objective Study of the Effects of Additional Training on Flying Skill, from Miller, N.E. (Editor), Psychological Research on Pilot Training, AAF Aviation Psychology Program Research Report No. 8, 1947.

Mixon, T.R., and Moroney, W.F., Annotated Bibliography of Objective Pilot Performance, Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN IH-330, March 1981. (In press).

Mood, A.M., Graybill, F.A., and Boes, D.C., Introduction to the Theory of Statistics, 3rd ed., McGraw-Hill, 1974.

Moore, S.B., Meshier, C.W., and Coward, R.E., The Good Stick Index. A Performance Measurement for Air Combat Training. First Interservice/Industry Training Equipment Conference, sponsored by Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN IH-316, November 1979, p. 145-154.

Morris, W.T., Management Science in Action, Richard D. Irwin, Inc., Homewood, IL, 1963.

Mundel, M.E., Motion and Time Study: Improving Productivity, Prentice-Hall, 1978.

Murphy, R., Individual Differences in Pilot Performance, Proceedings of the 6th Congress of the International Ergonomics Association, College Park, MD, July 1976, p. 403-409.

Norman, D.A., Memory and Attention, Wiley & Sons, Inc., 2nd ed., 1976.

North, R.A., and Griffin, G.R., Aviator Selection 1919-1977, Naval Aerospace Medical Research Laboratory, Pensacola, FL 32508, NAMRL-SR-77-2, October 1977, AD A048105.

Obermayer, R.W., and Muckler, F.A., Performance Measurement in Flight Simulation Studies, National Aeronautics and Space Administration, Washington, DC, NASA CR-82, July 1964.

Obermayer, R.W., and Vreuls, D., Combat-Ready Crew Performance Measurement System: Phase III A Crew Performance Measurement, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-74-108 (IV), December 1974, AD B005520.

Obermayer, R.W., Vreuls, D., Muckler, F.A., Conway, E.J., and Fitzgerald, J.A., Combat-Ready Crew Performance Measurement System: Final Report, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-74-108(I), December 1974, AD B005517.

Obermayer, R.W., Vreuls, D., Muckler, F.A., and Conway, E.J., Combat-Ready Crew Performance Measurement System: Phase III D Specifications and Implementation Plan, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-74-108 (VII), December 1974, AD B005522.

Office of the Secretary of Defense, Department of Defense Review of Tactical Jet Operational Readiness Training, Assistant Secretary of Defense (Manpower and Reserve Affairs), Washington, DC 20301, November 1968, AD 845259.

Oller, R.G., Human Factors Data Thesaurus (An Application to Task Data), Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH, AMRL-TR-67-211, March 1968, AD 670578.

Operations Committee, Naval Science Department, Naval Operations Analysis, U.S. Naval Institute, 1968.

Osborn, W.C., Developing Performance Tests for Training Evaluation, Human Resources Research Organization, Alexandria, VA 22314, HumRRO-PP-3-73, October 1971, AD 758436.

Parker, J.F. Jr., "The Identification of Performance Dimensions Through Factor Analysis," Human Factors, v. 9(4): p. 367-373, August 1967.

Parsons, H.M., Man-Machine System Experiments, Johns Hopkins Press, 1972.

Pickrel, E.W., and McDonald, T.A., "Quantification of Human Performance in Large, Complex Systems," Human Factors, v. 6(6): p. 647-662, December 1964.

Pierce, B.J., De Maio, J., Eddowes, E.E., and Yates, D., Airborne Performance Methodology Application and Validation: F-4 Pop-Up Training Evaluation, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-79-7, June 1979, AD A072611. See also Proceedings of the 23rd Annual Meeting of the Human Factors Society, p. 320-324, November 1979.

Prophet, W.W., Performance Measurement in Helicopter Training and Operations, Office of Chief of Research and Development (Army), Washington, DC 20310, HumRRO-PP-10-72, April 1972, AD 743157.

Prophet, W.W., Long-Term Retention of Flying S ills: A Review of the Literature, U.S. Air Force Headquarters (AF/SAA), Washington, DC 20332, HumRRO FR-ED(P)-76-35, October 1976, AD A036077.

Prophet, W.W., U.S. Navy Fleet Aviation Training Program Development, Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN 77-C-0009-1, March 1978, AD A055788.

Rankin, W.C., and McDaniel, W.C., Computer Aided Training Evaluation and Scheduling (CATES) System: Assessing Flight Task Proficiency, Training and Analysis Group, Orlando, FL 32813, TAEG Report No. 94, December 1980.

Read, J.H., Specification for Trainer, A-6E Weapon System Device 2F114, Grumman Aerospace Corporation, September 1974.

Ricketson, D.S., Pilot Error As A Cause of Army Helicopter Accidents, Proceedings of a Conference on Aircrew Performance in Army Aviation held at U.S. Army Aviation Center, Fort Rucker, AL, on November 27-29, 1973, Office of the Chief of Research, Development and Acquisition (Army) and U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA 22209, July 1974, p. 75-80, AD A001539.

Riis, E., Measurement of Performance in F-86K Simulator, Proceedings, Annual AGARD Meeting of the Aerospace Medical Panel, Toronto, Canada, Advisory Group for Aerospace Research and Development, Paris, France, AGARD Conference Proceedings No. 14, Assessment of Skill and Performance in Flying, September 1966, p. 49-56, AD 661165.

Rinsky, R., Pearlman, T., Nienaltowski, W., and Marchessault, C., Criteria Report for A-6E Weapon System Trainer Device 2F114, Final, Vol. II, Grumman Aerospace Corporation, TR-T01-CD-A005-04, January 1977.

Roscoe, S.N., Review of Flight Training Technology, U.S. Army Research Institute for the Behavioral and Social Sciences, Research Problem Review 76-3, July 1976, AD A076641.

Rosenmayer, C.E., and Asiala, C.F., F-18 Human Engineering Task Analysis Report, McDonnell Aircraft Company, St. Louis, MO 63166, MDC A4276, August 1976.

Ryack, B.L., and Krendel, E.S., Experimental Study of the Natural Pilot Flight Proficiency Evaluation Model, U.S. Naval Training Device Center, Port Washington, NY, NAVTRADEVCEN 323-2, April 1963, AD 414666.

Saleh, J., Thomas, J.O., and Boylan, R.J., Identification of Significant Aircrew Decisions in Navy Attack Aircraft, Naval Weapons Center, China Lake, CA 93555, NWC TP 6117, January 1980, AD B045034.

Sanders, M.G., Simmons, R.R., and Hofmann, M.A., "Visual Workload of the Copilot/Navigator During Terrain Flight," Human Factors, v. 21(3): p. 369-383, June 1979.

Schohan, B., Rawson, H.E., and Soliday, S.M., "Pilot and Observer Performance in Simulated Low Altitude High Speed Flight," Human Factors, v. 7(3): p. 257-265, June 1965.

Semple, C.A., Vreuls, D., Cotton, J.C., Durfee, D.R., Hooks, J.T., and Butler, E.A., Functional Design of an Automated Instructional Support System for Operational Flight Trainers, Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN 76-C-0096-1, January 1979, AD A065573.

Senders, J.W., A Study of the Attentional Demand of VFR Helicopter Pilotage, Proceedings of a Conference on Aircrew Performance in Army Aviation held at U.S. Army Aviation Center, Fort Rucker, AL on November 27-29, 1973, Office of the Chief of Research, Development and Acquisition (Army) and U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA 22209, July 1974, p. 175-187, AD A001539.

Shannon, R.H., and Waag, W.L., Toward the Development of a Criterion for Fleet Effectiveness in the F-4 Fighter Community, Proceedings of the 16th Annual Meeting of the Human Factors Society, Los Angeles, CA, October 1972, p. 335-337. See also NAMRL-1173, December 1972, AD 755184.

Shelnutt, J.B., Childs, J.M., Prophet, W.W., and Spears, W.D., Human Factors Problems in General Aviation, Federal Aviation Administration, Atlantic City, NJ 08405, FAA-CT-80-194, April 1980.

Shipley, B.D., Gerlach, V.S., and Brecke, F.H., Measurement of Flight Performance in a Flight Simulator, Air Force Office of Scientific Research, Bolling AFB, DC 20332, AFOSR-TR-75-0208, August 1974, AD A004488.

Shipley, B.D., An Automated Measurement Technique for Evaluating Pilot Skill, Air Force Office of Scientific Research, Bolling AFB, DC 20332, AFOSR-TR-76-1253, February 1976, AD A033920.

Siegel, A.I., and Federman, P.J., Increasing ASW Helicopter Effectiveness Through Communications Training, Naval Training Device Center, Orlando, FL, NAVTRADEVCEN 66-C-0045-1, October 1968, AD 682498.

Siegel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, 1956.

Singleton, W.T., Fox, J.G., and Whitfield, D., Measurement of Man At Work, Taylor & Francis LTD, 1971.

Smit, J., Pilot Workload Analysis Based Upon In-Flight Physiological Measurements and Task Analysis Methods, National Aerospace Laboratory (NLR), The Netherlands, NLR-MP-76001-U, June 1976, AD B026957.

Smith, B.A., Development and Inflight Testing of a Multi-Media Course for Instructing Navigation for Night Nap-of-the-Earth Flight, Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, CA, October 1980, p. 568-571.

Smode, A.F., Gruber, A., and Ely, J.H., The Measurement of Advanced Flight Vehicle Crew Proficiency in Synthetic Ground Environments, 6570th Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH, MRL-TDR-62-2, February 1962, AD 273449.

Smode, A.F., and Meyer, D.E., Research Data and Information Relevant to Pilot Training. Volume I. General Features of Air Force Pilot Training and Some Research Issues, Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH, AMRL-TR-66-99-Vol I, July 1966, AD 801776.

Soliday, S.M., "Navigation in Terrain-Following Flight," Human Factors, v. 12(5): p. 425-433, October 1970.

Steyn, D.W., "The Criterion - Stagnation or Development? A Selective Review of Published Research Work on Criterion Problems," Psychologia Africana, v. 12(3): p. 193-211, March 1969.

Swezey, R.W., "Aspects of Criterion-Referenced Measurement in Performance Evaluation," Human Factors, v. 20(2): p. 169-178, April 1978.

Swink, J.R., Butler, E.A., Lankford, H.E., Miller, R.M., Watkins, H., and Waag, W.L., Definition of Requirements for Performance Measurement System for C-5 Aircrew Members, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-78-54, October 1978, AD A063282.

Thorndike, R.L., Reliability, from Lindquist, E.F. (Editor), Educational Measurement, George Banta Publishing Company, 1951.

Tindle, J.R., Effects on A-6E Bombardier/Navigator Flight Training with the Introduction of Device 2F114, A-6E Weapon System Trainer, Master's Thesis, Naval Postgraduate School, March 1979.

Uhlaner, J.E., and Drucker, A.J., "Criteria for Human Performance Research," Human Factors, v. 6(3): p. 265-278, June 1964.

Uhlaner, J.E., and Drucker, A.J., "Military Research on Performance Criteria: A Change of Emphasis," Human Factors, v. 22(2): p. 131-139, April 1980.

Van Cott, H.P., and Kinkade, R.G. (Editors), Human Engineering Guide to Equipment Design, Joint Army-Navy-Air Force Steering Committee, 1972.

Voiers, W.D., A Comparison of the Components of Simulated Radar Bombing Error in Terms of Reliability and Sensitivity to Practice, Air Force Personnel & Training Research Center, Lackland AFB, TX, AFPTRC-54-74, April 1954, AD 62566.

Vreuls, D., and Obermayer, R.W., Study of Crew Performance Measurement for High-Performance Aircraft Weapon System Training: Air-to-Air Intercept, Naval Training Device Center, Orlando, FL 32813, NAVTRADEVCEN 70-C-0059-1, February 1971, AD 727739.

Vreuls, D., and Obermayer, R.W., Emerging Developments in Flight Training Performance Measurement, from U.S. Naval Training Device Center 25th Anniversary Commemorative Technical Journal, Orlando, FL 32813, November 1971, p. 199-210.

Vreuls, D., Obermayer, R.W., Goldstein, I., and Lauber, J.K., Measurement of Trainee Performance in a Captive Rotary-Wing Device, Naval Training Equipment Center, Orlando, FL 32813, June 1973, AD 764088.

Vreuls, D., Obermayer, R.W., and Goldstein, I., Trainee performance Measurement Development Using Multivariate Measure Selection Techniques, Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN 73-C-0066-1, September 1974, AD 787594. See also Proceedings, NTEC/Industry Conference, Orlando, FL, NAVTRAEQUIPCEN IH-240, November 1974, P. 227-236, AD A000970.

Vreuls, D., Wooldridge, A.L., Obermayer, R.W., Johnson, R.M., Norman, D.A., and Goldstein, I., Development and Evaluation of Trainee Performance Measures in an Automated Instrument Flight Maneuvers Trainer, Naval Training Equipment Center, Orlando, FL 32813, NAVTRAEQUIPCEN 74-C-0063-1, October 1975, AD A024517.

Vreuls, D., and Wooldridge, A.L., Aircrew Performance Measurement, Proceedings of the Symposium on Productivity Enhancement: Personnel Assessment in Navy Systems, Naval Personnel Research and Development Center, San Diego, CA, October 1977.

Vreuls, D., and Cotton, J.C., Feasibility of Aircrew Performance Measurement Standardization for Research, Air Force Office of Scientific Research, Bolling AFB, DC 20332, April 1980.

Waag, W.L., Eddowes, E.E., Fuller, J.H. Jr., and Fuller, R.R., ASUPT Automated Objective Performance Measurement System, Air Force Human Resources Laboratory, Williams AFB, AZ 85224, AFHRL-TR-75-3, March 1975, AD A014799.

Welford, A.T., Fundamentals of Skill, Methuen & Co. LTD, 1971.

Welford, A.T., Skilled Performance: Perceptual and Motor Skills, Scott, Foresman and Company, 1976.

Williams, A.C. Jr., Simon, C.W., Haugen, R., and Roscoe, S.N., Operator Performance in Strike Reconnaissance, Wright Air Development Division, Wright-Patterson AFB, OH, WADD Technical Report 60-521, August 1960, AD 246545.

APPENDIX A

A-6E TRAM RADAR NAVIGATION TASK LISTING

The enclosed task listing for the B/N during radar navigation in the A-6E WST was compiled from various sources as discussed in Chapter V. This was the first phase of developing a task analysis for the purpose of measuring performance of the B/N during radar navigation.

A-6E TRAM RADAR NAVIGATION TASK LISTING

SEGMENT 1: AFTER TAKEOFF CHECKS

T1 ADJUST RADAR PANEL FOR SCOPE RETURN

S1 SET SEARCH RADAR PWR SWITCH..................ON

S2 SET XMT SWITCH..............................NORM

T2 ACTIVATE SYSTEM STEERING TO INITIAL POINT (IP)

S1 CHECK COMPTMODE SWITCH....................STEER

S2 DEPRESS TGT N ADDRESS KEY HAVING IP LAT/LONG

S3 CHECK COMPT/MAN SWITCH....................COMPT

T3 CHECK FOR ACCURATE SYSTEM STEERING TO IP

S1 READ SYSTEM BEARING AND RANGE TO IP FROM DVRI BUG AND RANGE DISPLAYS

S2 COMPARE SYSTEM BEARING AND RANGE TO IP WITH PRE-PLANNED OR ESTIMATED BEARING AND RANGE TO IP

S3 GO TO T5 IF SYSTEM STEERING TO IP IS CORRECT

S4 GO TO T4 IF SYSTEM STEERING TO IP IS NOT CORRECT

T4 TROUBLESHOOT SYSTEM STEERING IF REQUIRED

S1 DETERMINE IP LAT/LONG FROM CHART OR IFR SUPPLEMENT

S2 COMPARE SYSTEM IP LAT/LONG WITH ACTUAL IP LAT/LONG

(a) THROW DDU DATA SWITCH..............ON CALL

(b) READ SYSTEM TGT N ADDRESS (IP) FROM LOWER DDU LAT/LONG DISPLAYS

S3 INSERT CORRECT IP LAT/LONG IF REQUIRED

S4 EVALUATE SYSTEM PRESENT POSITION AND ESTIMATED POSITION FROM CHART

(a) THROW DDU DATA SWITCH.............PRES POS

(b) READ SYSTEM PRESENT POSITION LAT/LONG FROM LOWER DDU LAT/LONG DISPLAYS

(c) COMPARE SYSTEM PRESENT POSITION WITH ESTIMATED PRESENT POSITION FROM CHART

(d) INSERT CORRECT PRESENT POSITION IF REQUIRED

(1) DEPRESS PRES LOC ADDRESS KEY ON COMPUTER KEYBOARD

(2) THROW COMPTMODE SWITCH..........ENTER

(3) INSERT CORRECT PRESENT POSITION LAT/LONG

(4) THROW COMPTMODE SWITCH..........STEER

(5) CHECK LOWER DDU LAT/LONG DISPLAYS FOR ACCURATE DATA ENTRY

S5 INFORM PILOT OF APPROXIMATE HEADING TO IP IF REQUIRED AND GO TO T6

T5 INFORM PILOT THAT SYSTEM STEERING IS TO IP

S1 CHECK THAT PILOT MAINTAINS SAFE FLIGHT AND FOLLOWS SYSTEM STEERING

T6 CHECKOUT RADAR FOR STATUS

S1 GO TO NEXT EVENT IF RADAR RETURN IS PRESENT

S2 GO TO T7 IF RADAR RETURN IS NOT PRESENT

T7 TROUBLESHOOT RADAR IF REQUIRED

S1 CHECK TEST MODE SWITCH (RADAR/DRS TEST).................................. CENTERED

S2 CHECK FAULT ISLN SWITCH................CENTERED

S3 CHECK RADAR CIRCUIT BREAKER..................IN

S4 USE PCL CHECKLIST FOR RADAR TURN-UP PROCEDURES

S5 ALERT PILOT IF RADAR INOPERATIVE AND ABORT RADAR NAVIGATION FLIGHT

SEGMENT 2: NAVIGATION TO IP

T1 TUNE RADAR FOR OPTIMUM PPI DISPLAY

S1 ROTATE CONTRAST CONTROL......................CW

S2 ROTATE BRT CONTROL (UNTIL SWEEP PRESENT).....CW

S3 CHECK VIDEO/DIF CONTROLS....................CCW

S4 ROTATE RCVR CONTROL (UNTIL RETURN IS PRESENT)....................CW

S5 CHECK DISPLAYS BUTTONS......................PPI

S6 ADJUST PPI RANGE CONTROL (UNTIL IP AT TOP OF SCOPE)...............CW/CCW

S7 ROTATE RNG MKR/AZ MKR CONTROLS (UNTIL CURSORS PRESENT)......................CW

S8 CHECK SCAN STAB CONTROL.....................ADL

S9 ROTATE SCAN ANGLE CONTROL (UNTIL DESIRED SWEEP WIDTH PRESENT)......CW/CCW

S10 CHECK SCAN RATE SWITCH.....................FAST

S11 CHECK AMTI CONTROL..........................CCW

S12 CHECK STC SLOPE/DEPTH CONTROLS...........CW/CCW

S13 THROW ANT PATT SWITCH (CONTINUE FOR 10 SECOND MINIMUM)............FAR

S14 CHECK RCVR SWITCH...........................AFC

S15 CHECK BEACON CONTROL........................CCW

S16 CHECK AZ-RNG TRKG SWITCH...................OFF

S17 CHECK <u>FREQ AGILITY</u> SWITCH....................ON

T2 POSITION CURSOR INTERSECTION ON IP IF REQUIRED

S1 PLACE LEFT HAND ON SLEW CONTROL STICK

S2 DEPRESS RADAR SLEW BUTTON AND PUSH SLEW CONTROL STICK IN DIRECTION OF DESIRED CURSOR INTERSECTION MOVEMENT

S3 COMPARE RADAR RETURN IMAGE ON CURSOR INTERSECTION WITH PREPLANNED CHART AND QUICK & DIRTY IP

S4 REPEAT S1 AND S2 IF REQUIRED

S5 PUSH <u>CORRECT POS</u> BUTTON ON LOWER DDU PANEL

T3 ACTIVATE DOPPLER RADAR

S1 SET <u>DOPPLER CONTROL</u> SWITCH (LAND OR SEA AS APPROPRIATE).................ON

S2 MONITOR DOPPLER CONTROL PANEL FOR DOPPLER RADAR STATUS

(a) OBSERVE MEMORY LIGHT OUT FOR PROPER OPERATION

(b) OBSERVE DRIFT AND GND SPEED DISPLAYS FOR DRIFT ANGLE AND GROUND SPEED PRESENT

T4 SELECT SYSTEM STEERING OR DEAD RECKONING (DR) NAVIGATION

S1 DETERMINE STATUS OF INS, RADAR, AND DOPPLER RADAR

S2 OBSERVE CURSOR INTERSECTION DRIFT ON RADAR SCOPE

S3 GO TO T5 IF DEAD RECKONING IS SELECTED (LARGE CURSOR INTERSECTION DRIFT) OR SEGMENT 3, T1 IF SYSTEM NAVIGATION IS SELECTED (SMALL CURSOR INTERSECTION DRIFT)

T5 PERFORM DEAD RECKONING NAVIGATION (BEYOND THE SCOPE OF THIS REPORT)

SEGMENT 3: NAVIGATION TO TURN POINT (TP)

T1 INITIATE TURN AT IP

S1 ALERT PILOT OF NEXT OUTBOUND HEADING ONE MINUTE PRIOR TO REACHING IP

S2 SET OUTBOUND HEADING ON HORIZONTAL SITUATION INDICATOR (HSI) BY ROTATING HEADING SELECT KNOB UNTIL THE HEADING SELECT MARKER IS ON THE DESIRED HEADING

S3 CHECK AZ-RNG TRKG SWITCH....................OFF

S4 MONITOR DVRI HEADING BUG FOR MOVEMENT TO 180° RELATIVE POSITION (IP PASSAGE)

S5 ACTIVATE COCKPIT CLOCK AT IP PASSAGE

S6 INFORM PILOT OF IP PASSAGE AND OUTBOUND HEADING

S7 CHECK FOR PILOT TURNING TO NEW HEADING

T2 ACTIVATE SYSTEM STEERING TO TP

S1 DEPRESS TGT N ADDRESS KEY HAVING NEXT TP LAT/LONG

S2 CHECK COMPTMODE SWITCH...................STEER

T3 CHECK FOR ACCURATE SYSTEM STEERING TO TP

S1 READ SYSTEM BEARING AND RANGE TO TP FROM DVRI BUG AND RANGE DISPLAYS

S2 COMPARE SYSTEM BEARING AND RANGE TO TP WITH PREPLANNED OR ESTIMATED BEARING AND RANGE TO TP

S3 GO TO T5 IF SYSTEM STEERING TO TP IS CORRECT

S4 GO TO T4 IF SYSTEM STEERING TO TP IS NOT CORRECT

T4 TROUBLESHOOT SYSTEM STEERING IF REQUIRED

S1 DETERMINE TP LAT/LONG FROM CHART OR IFR SUPPLEMENT

S2 COMPARE SYSTEM TP LAT/LONG WITH ACTUAL TP LAT/LONG

(a) THROW <u>DDU DATA</u> SWITCH.............ON CALL

(b) READ SYSTEM <u>TGT N</u> ADDRESS (TP) FROM LOWER DDU LAT/LONG DISPLAYS

S3 INSERT CORRECT TP LAT/LONG IF REQUIRED

S4 EVALUATE SYSTEM PRESENT POSITION AND ESTIMATED POSITION FROM CHART

(a) THROW <u>DDU DATA</u> SWITCH............PRES POS

(b) READ SYSTEM PRESENT POSITION LAT/LONG FROM LOWER DDU LAT/LONG DISPLAYS

(c) COMPARE SYSTEM PRESENT POSITION WITH ESTIMATED PRESENT POSITION FROM CHART

(d) INSERT CORRECT PRESENT POSITION IF REQUIRED

(1) DEPRESS <u>PRES LOC</u> ADDRESS KEY ON COMPUTER KEYBOARD

(2) THROW <u>COMPTMODE</u> SWITCH.........ENTER

(3) INSERT CORRECT PRESENT POSITION LAT/LONG

(4) THROW <u>COMPTMODE</u> SWITCH.........STEER

(5) CHECK LOWER DDU LAT/LONG DISPLAYS FOR ACCURATE DATA ENTRY

T5 INFORM PILOT THAT SYSTEM STEERING IS TO THE TP

S1 CHECK THAT PILOT MAINTAINS SAFE FLIGHT AND FOLLOWS SYSTEM STEERING

T6 INSERT DATA FOR NEXT REMAINING TPs IF REQUIRED

S1 THROW <u>COMPTMODE</u> SWITCH...................ENTER

S2 DEPRESS PREVIOUSLY UTILIZED <u>TGT N</u> ADDRESS KEY

S3 DEPRESS <u>POS</u> ACTION KEY

S4 DEPRESS APPROPRIATE QUANTITY KEYS IN SEQUENCE FOR TP LAT/LONG

S5 DEPRESS ALT ACTION KEY

S6 DEPRESS APPROPRIATE QUANTITY KEYS IN SEQUENCE FOR TP ALTITUDE

S7 ALERT PILOT TO MAINTAIN PRESENT HEADING

S8 THROW COMPTMODE SWITCH....................STEER

S9 THROW DDU DATA SWITCH...................ON CALL

S10 CHECK LOWER DDU LAT/LONG DISPLAYS FOR ACCURATE DATA ENTRY

S11 REPEAT S1 THROUGH S10 IF REQUIRED

S12 DEPRESS REQUIRED TGT N ADDRESS KEY FOR CURRENT TP SYSTEM STEERING

S13 CHECK FOR ACCURATE SYSTEM STEERING TO TP

(a) READ SYSTEM BEARING AND RANGE TO TP FROM DVRI BUG AND RANGE DISPLAYS

(b) COMPARE SYSTEM BEARING AND RANGE TO TP WITH PREPLANNED OR ESTIMATED BEARING AND RANGE TO TP

S14 REPEAT S12 AND S13 IF REQUIRED

S15 INFORM PILOT THAT SYSTEM STEERING IS TO THE TP

(a) CHECK THAT PILOT MAINTAINS SAFE FLIGHT AND FOLLOWS SYSTEM STEERING

T7 PERFORM SYSTEM NAVIGATION TASKS

S1 TUNE RADAR FOR OPTIMUM PPI DISPLAY

(a) ADJUST PPI RANGE CONTROL (UNTIL TP AT TOP OF SCOPE)..............CW

(b) ADJUST RCVR CONTROL (UNTIL RETURN IS ENHANCED)..........CW/CCW

(c) ADJUST STC DEPTH CONTROL (UNTIL EVEN RETURN PRESENT).............CW

(d) ADJUST SCAN ANGLE CONTROL (UNTIL DESIRED SWEEP WIDTH PRESENT)..CW/CCW

S2 MONITOR FLIGHT PROGRESS USING RADAR SIGNIFICANT TERRAIN/CULTURAL FEATURES AS CHECK POINTS

(a) COMPARE RATAR RETURN IMAGE WITH PREPLANNED CHART AND QUICK & DIRTY

(b) COMPARE SYSTEM PRESENT POSITION WITH ESTIMATED PRESENT POSITION FROM CHART IF REQUIRED

(1) CHECK DDU DATA SWITCH.........PRES POS

(2) READ SYSTEM PRESENT POSITION LAT/LONG FROM LOWER DDU LAT/LONG DISPLAYS

(3) RECORD SYSTEM PRESENT POSITION ON CHART

(4) READ TIME-INTO-LEG OR TOTAL TIME FROM COCKPIT CLOCK

(5) RECORD ESTIMATED PRESENT POSITION ON CHART

(c) REPEAT (a) AND (b) IF REQUIRED

S3 INFORM PILOT OF SYSTEM NAVIGATIONAL ACCURACY

S4 POSITION CURSOR INTERSECTION ON TP IF REQUIRED

(a) PLACE LEFT HAND ON SLEW CONTROL STICK

(b) DEPRESS RADAR SLEW BUTTON AND PUSH SLEW CONTROL STICK IN DIRECTION OF DESIRED CURSOR INTERSECTION MOVEMENT

(c) COMPARE RADAR RETURN IMAGE ON CURSOR INTERSECTION WITH PREPLANNED CHART AND QUICK & DIRTY TP

(d) REPEAT (a) AND (b) IF REQUIRED

(e) PUSH CORRECT POS BUTTON ON LOWER DDU PANEL

S5 MONITOR NAVIGATIONAL EQUIPMENT FOR OPERATING STATUS OR EQUIPMENT CONDITION

(a) CHECK NAVIGATION (INS) CONTROL PANEL FOR INS FAILURE INDICATIONS

(b) CHECK VTR PANEL FOR OPERATING INDICATIONS IF REQUIRED

(c) CHECK FOR COMPUTER ERROR LIGHT ON DVRI PANEL AND LOWER DDU PANEL

(d) CHECK ATTITUDE REF SWITCH..........COMP IN

(e) CHECK MAGNETIC VARIATION FROM MAG VAR DISPLAY ON LOWER DDU PANEL

(1) COMPARE TO CHART MAGNETIC VARIATION AND SET IF NECESSARY

(f) INFORM PILOT OF NAVIGATIONAL EQUIPMENT STATUS

S6 MONITOR TIME ON TP

(a) COMPARE RECORDED ACTUAL LEG OR TOTAL TIME FOR PREVIOUS TP WITH PREPLANNED LEG OR TOTAL TIME FOR PREVIOUS TP

(b) INSTRUCT PILOT TO ADJUST THROTTLE CONTROLS SO AS TO CORRECT TIME ON TP

(c) INFORM PILOT OF TIME ON TP RESULTS

S7 MONITOR SYSTEM VELOCITIES

(a) READ SYSTEM GROUND SPEED (GS) AND WIND

(1) THROW DDU DATA SWITCH.............DATA

(2) RECORD GROUND SPEED IN G/S DISPLAY

(3) RECORD WIND SPEED AND DIRECTION IN WIND SPEED/WIND DIR DISPLAYS

(b) READ AND RECORD GS FROM DOPPLER PANEL GND SPEED DISPLAY

(c) READ AND RECORD INDICATED AIRSPEED (IAS) FROM AIRSPEED INDICATOR ON PILOT'S INSTRUMENT PANEL

(d) READ SYSTEM TRUE AIRSPEED (TAS)

(1) THROW DDU DATA SELECT SWITCH........A

(2) RECORD TAS FROM DISPLAY 1

(e) EVALUATE SYSTEM VELOCITIES USING GS, WIND, IAS, AND TAS

(f) TROUBLESHOOT SYSTEM VELOCITIES IF REQUIRED

(g) INFORM PILOT OF SYSTEM VELOCITY RESULTS

S8 MONITOR SYSTEM HEADING

(a) READ TRUE HEADING FROM DVRI DISPLAY BUG

(b) READ MAGNETIC HEADING FROM WET COMPASS

(c) READ HEADING FROM HORIZONTAL SITUATION INDICATOR

(d) READ MAGNETIC VARIATION FROM MAG VAR DISPLAY ON LOWER DDU PANEL

(e) EVALUATE HEADING ACCURACY USING TRUE HEADING, MAGNETIC VARIATION, AND COMPASS HEADING DATA WITH "CDMVT" FORMULA

(f) ADJUST MA-1 COMPASS NEEDLE DEFLECTION WITH PULL TO SET CONTROL IF REQUIRED

(g) TROUBLESHOOT SYSTEM HEADING IF REQUIRED

(h) INSTRUCT PILOT TO ADJUST HEADING SO AS TO MAINTAIN PREPLANNED COURSE

(i) INFORM PILOT OF SYSTEM HEADING RESULTS

S9 MONITOR SYSTEM ALTITUDE

(a) READ ALTITUDE (AGL) FROM RADAR ALTIMETER

(b) READ PRESSURE ALTITUDE (MSL) FROM PRESSURE ALTIMETER

(c) READ SYSTEM ALTITUDE (MSL)

(1) THROW DDU DATA SWITCH.........PRES POS

(2) READ ALTITUDE IN ALT DISPLAY

(d) READ TERRAIN ALTITUDE FOR PRESENT POSITION ON CHART

(e) EVALUATE SYSTEM ALTITUDE ACCURACY USING ABOVE SOURCES

(f) TROUBLESHOOT SYSTEM ALTITUDE IF REQUIRED

(g) INSERT CORRECT ALTITUDE IF REQUIRED

(h) INFORM PILOT OF SYSTEM ALTITUDE RESULTS

S10 MONITOR SAFETY OF FLIGHT INSTRUMENTS AND EQUIPMENT

(a) EVALUATE FUEL STATUS BY COMPARING ACTUAL FUEL REMAINING WITH PREPLANNED FUEL REMAINING

(b) INFORM PILOT OF FUEL STATUS

(c) CHECK ANNUNCIATOR CAUTION LIGHTS FOR EMERGENCY INDICATIONS

(d) CHECK FIRE WARNING LIGHTS FOR AIRCRAFT FIRE INDICATIONS

(e) CHECK ACCESSIBLE CIRCUIT BREAKERS........IN

(f) INFORM PILOT OF SAFETY OF FLIGHT INSTRUMENTS AND EQUIPMENT RESULTS

T8 PERFORM APPROACH TO TP PROCEDURES

S1 CHECK FLIGHT PROGRESS USING RADAR SIGNIFICANT TERRAIN/CULTURAL FEATURES AS CHECK POINTS

(a) COMPARE RADAR RETURN IMAGE ON CURSOR INTERSECTION WITH PREPLANNED CHART AND QUICK & DIRTY

(b) POSITION CURSOR INTERSECTION ON TP IF REQUIRED

(c) REPEAT (a) AND (b) IF REQUIRED

(d) PUSH CORRECT POS BUTTON ON LOWER DDU PANEL

S2 SELECT ARE 30/60 DISPLAY AT APPROXIMATELY 17 MILES FROM TP AND NAVIGATION IS ACCURATE

S3 OBSERVE ARE 30/60 EXPANDED DISPLAY AT APPROXIMATELY 17 MILES

S4 TUNE RADAR FOR OPTIMUM ARE 30/60 DISPLAY

(a) ROTATE STC SLOPE CONTROL................CCW

(b) THROW ANT PATT SWITCH (CONTINUE UNTIL BRIGHTEST RETURN PRESENT)........NEAR

(c) ADJUST RCVR CONTROL (UNTIL RETURN IS ENHANCED)..............CCW

(d) ADJUST SCAN ANGLE CONTROL............CW/CCW

(e) CHECK AZ-RNG TRKG SWITCH................OFF

(f) CHECK ELEV TRKG SWITCH..................OFF

(g) ADJUST VIDEO/DIF CONTROLS TO ENHANCE RETURN RESOLUTION IF REQUIRED

S5 CONTINUE POSITIONING CURSOR INTERSECTION ON FRONT LEADING EDGE CENTER OF TURN POINT RETURN

T9 PERFORM VELOCITY CORRECT PROCEDURES

S1 GO TO T9.1 FOR AUTOMATIC VELOCITY CORRECT (AZ-RANGE LOCK-ON OR TRACK-WHILE-SCAN REQUIRED)

S2 GO TO T9.2 FOR MANUAL VELOCITY CORRECT

S3 FLIR TRACKING MANUAL VELOCITY CORRECT (NOT PART OF SIMULATOR CAPABILITY)

T9.1 PERFORM AUTOMATIC VELOCITY CORRECT

S1 POSITION AZIMUTH CURSOR TO CENTER OF TP RETURN

S2 POSITION RANGE CURSOR TO JUST LEADING EDGE OF TP RETURN

S3 THROW AZ-RNG TRKG SWITCH....................ON

S4 CHECK AZ-RANGE INDICATOR LIGHT...............ON

S5 ROTATE VELOCITY CORRECT SWITCH.....MEMORY POINT

S6 CHECK DDU DATA SELECT SWITCH..................A

S7 CHECK FLT DATA DISPLAY 4 FOR SOME VALUE GREATER THAN 000

S8 ROTATE VELOCITY CORRECT SWITCH (BEFORE TP WALK-DOWN)..................OFF SAVE

T9.2 PERFORM MANUAL VELOCITY CORRECT

S1 POSITION AZIMUTH CURSOR TO CENTER OF TP RETURN

S2 POSITION RANGE CURSOR TO JUST LEADING EDGE OF TP RETURN

S3 CHECK AZ-RNG TRKG SWITCH....................OFF

S4 ROTATE VELOCITY CORRECT SWITCH.....MEMORY POINT

S5 DELAY FOR 10 SECOND MINIMUM/128 SECOND MAXIMUM TO ALLOW CURSOR DRIFT

S6 REPEAT POSITIONING BEARING AND RANGE CURSORS TO JUST LEADING EDGE CENTER OF TP RETURN

S7 MONITOR FURTHER CURSOR DRIFT AND REPEAT POSITIONING IF REQUIRED

S8 CHECK DDU DATA SELECT SWITCH..................A

S9 CHECK FLT DATA DISPLAY 4 FOR SOME VALUE GREATER THAN 000

S10 ROTATE VELOCITY CORRECT SWITCH (BEFORE TP WALK-DOWN)..................OFF SAVE

T10 INITIATE TURN AT TP

S1 ALERT PILOT OF NEXT OUTBOUND HEADING ONE MINUTE PRIOR TO REACHING TP

S2 SET OUTBOUND HEADING ON HSI

S3 CHECK AZ-RNG TRKG SWITCH....................OFF

S4 MONITOR DVRI HEADING BUG FOR MOVEMENT TO 180° RELATIVE POSITION (TP PASSAGE)

S5 RECORD LEG TIME OR TOTAL ELAPSED TIME AT TP PASSAGE

S6 ACTIVATE COCKPIT CLOCK AT TP PASSAGE IF REQUIRED (LEG TIME ONLY)

S7 INFORM PILOT OF TP PASSAGE AND OUTBOUND HEADING

S8 CHECK FOR PILOT TURNING TO NEW HEADING

For subsequent navigation legs, return to Segment 3, Task 2

APPENDIX B

GLOSSARY OF TASK SPECIFIC BEHAVIORS

The enclosed glossary of 31 specific behaviors, or action verbs, was excerpted from Oller [1968]. Each verb has a specific meaning and is acceptable in the sense that all synonyms have been eliminated. Each action verb is used in the task listing (Appendix A), task analysis (Appendix C), and the MTLA (Appendix D), for the purpose of defining observable behavior that may be measured in terms of task performance.

Activate - Provide the initial force or action to begin an operation of some equipment configuration.

Adjust - Manipulate controls, levers, linkages and other equipment items to return equipment from an out-of-tolerance condition to an in-tolerance condition.

Alert - Inform designated persons that a certain condition exists in order to bring them up to a watchful state in which a quick reaction is possible.

Check - Examine to determine if a given action produces a specified result; to determine that a presupposed condition actually exists, or to confirm or determine measurements by the use of visual, auditory, tactile, or mechanical means.

Checkout - Perform routine procedures, which are discrete, ordered stepwise actions designed to determine the status or assess the performance of an item.

Compare - Examine the characteristics of two or more items to determine their similarities and differences.

Continue - Proceed in the performance of some action, procedure, etc., or to remain on the same course or direction (e.g., continue to check the temperature fluctuations; continue to adjust the controls; and continue on the same heading).

Delay - Wait a brief period of time before taking a certain action or making a response.

Depress - Apply manual (as opposed to automatic) pressure to activate or initiate an action or to cause an item of equipment to function or cease to function.

Determine - Find, discover, or detect a condition (e.g., determine degree of angle).

Evaluate - Judge or appraise the worth or amount, of a unit of equipment, operational procedure or condition (e.g., evaluate status of life support systems).

Inform - Pass on information in some appropriate manner to one or more persons about a condition, event, etc., that they should be aware of.

Initiate - Give a start to a plan, idea, request, or some form of human action (e.g., initiate a new safety procedure).

Insert - Place, put, or thrust something within an existing context (e.g., insert a part in the equipment, insert a request in the computer).

Instruct - Impart information in an organized, systematic manner to one or more persons.

Monitor - Observe continually or periodically visual displays, or listen for or to audio displays, or vibrations in order to determine equipment condition or operating status.

Observe - Note the presence of mechanical motion, the condition of an indicator, or audio display, or other sources of movement or audible sounds on a nonperiodic basis.

Perform - Carry out some action from preparation to completion (It is understood that some special skill or knowledge is required to successfully accomplish the action.).

Place - Transport an object to an exact location.

Position - Turn, slide, rotate, or otherwise move a switch, lever, valve handle, or similar control device to a selected orientation about some fixed reference point.

Push - Exert a force on an object in such a manner that the object will move or tend to move away from the origin of the force.

Read - Use ones eyes to comprehend some standardized form of visual symbols (e.g., sign, gauge, or chart).

Record - Make a permanent account of the results of some action, test, event, etc., so that the authentic evidence will be available for subsequent examination.

Repeat - Perform the same series of tests, operations, etc., over again, or perform an identical series of tasks, tests, operations, etc.

Rotate - Apply manual torque to cause a multiple position rotary switch or a constantly varying device like a handwheel, thumbwheel, or potentiometer to move in a clockwise or counter-clockwise manner.

Select - Choose, or to be commanded to choose, an alternative from among a series of similar choices (e.g., select a proper transmission frequency).

Set - Move pointers, clock hands, etc., to a position in conformity with a standard, or place mechanical controls in a predetermined position.

Throw - Change manually the setting of a toggle switch from one position to another.

Troubleshoot - Examine and analyze failure reports, equipment readouts, test equipment meter valves, failure symptoms, etc., to isolate the source of malfunction.

Tune - Adjust an item of equipment to a prescribed operating condition.

Use - Utilize some unit of equipment or operational procedure.

APPENDIX C

A-6E TRAM RADAR NAVIGATION TASK ANALYSIS

The purpose of performing a task analysis for measuring B/N performance during radar navigation was to provide candidate performance measure metrics that may describe either successful task performance or B/N skill acquisition. Only segment three, Navigation to TP, was examined to limit the scope of the task analysis. The sequential flow of tasks, subtasks, and subtask elements (defined below), is the same as that found in the A-6E TRAM radar navigation task listing (Appendix A). The seven columns of the task analysis form were defined in Chapter V but the definitions are repeated here for the convenience of the reader:

(1) Subtask - a component activity of a task. Within a task, collectively all subtasks comprise the task. Subtasks are represented by the letter "S" followed immediately by a numeral. Subtask elements are represented by a small letter in parentheses.

(2) Feedback - the indication of adequacy of response or action. Listed as VISUAL, TACTILE, AUDITORY, or VESTIBULAR and is listed in the subtask column for convenience only.

(3) Action Stimulus - the event or cue that instigates performance of the subtask. This stimulus may be an out-of-tolerance display indication, a requirement of periodic inspection, a command, a failure, etc.

(4) Time - the estimated time in seconds to perform the subtask or task element calculated from initiation to completion.

(5) Criticality - the relationship between mission success and the below-minimum performance or required excessive performance time of a particular subtask or subtask element. "High" (H) indicates poor subtask performance may lead to mission failure or an accident. "Medium" (M) indicates the possibility of degraded mission capability. "Low" (L) indicates that poor performance may have little effect on mission success.

(6) Potential Error - errors are classified as failure to perform the task (OMIT), performing the task inappropriately in time or accuracy (COMMIT), or performing sequential task steps in the incorrect order (SEQUENTIAL).

(7) Skills Required - the taxonomy of training objectives used for the Grumman task analysis was retained and presented in Table VII [Campbell, et al., 1977].

(8) Performance Measure Metrics - a candidate metric which may best describe the successful performance of the task or a genuine display of the required skills. The types of metrics suggested were classified as: TIME (time in seconds from start to finish of task), T-S (time-sharing or proportion of time that particular task is performed in relation to other tasks being performed in the same time period), R-T (reaction time in seconds from the onset of an action stimulus to task initiation), ACC (accuracy of task performance), FREQ (number of task occurrences), DEC (decisions made as a correct or incorrect choice depending on the particular situation and mission requirements), QUAL (quality of a task, especially in regards to radar scope tuning quality), and SUBJ (subjective observation or comprehension of the task execution success by an instructor).

The right-hand column of the task analysis form ("performance measure metrics") provided several hundred possible candidate measures for describing successful task performance or B/N skill acquisition. Using initial measure selection criteria as outlined in Chapter IV, these measures were reduced and combined with literature review candidate measures (Table II of Chapter IV) to produce the final candidate measure set as shown in Table XI of Chapter VII.

SEG 3: NAVIGATION TO TP | TASK: T1 INITIATE TURN AT IP

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRICS
S1 Alert pilot of next outbound heading one minute prior to reaching IP. Feedback: Auditory	Directive	05	M	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S2 Set outbound heading on HSI by rotating heading select knob until the heading select marker is on the desired heading. Feedback: Visual	Directive	05	L	Omit	Complex Procedure	TIME, T-S, ACC, FREQ
S3 Check AZ-RNG TRKG switch ...off. Feedback: Visual	Directive	02	L	Omit	Simple Procedure	ACC, FREQ, SUBJ
S4 Monitor DVRI heading bug for movement to 180° relative position (IP passage). Feedback: Visual	Mission	38	H	Omit	Complex Procedure	R-T, FREQ
S5 Activate cockpit clock at IP passage. Feedback: Tactile and Visual	Sequential	02	H	Omit	Simple Procedure	R-T, FREQ

SEG 3: NAVIGATION TO TP	TASK: T1 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRICS
S6 Inform pilot of IP passage and outbound heading. Feedback: Auditory	Sequential	05	M	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S7 Check for pilot turning to new heading. Feedback: Vestibular and Visual	Sequential	30	H	Omit	Simple Procedure	ACC

SEG 3: NAVIGATION TO TP	TASK: T2 ACTIVATE SYSTEM STEERING TO TP					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRICS
S1 Depress TGT N address key having next TP Lat/Long. Feedback: Tactile and Visual	Mission	02	H	Sequence	Complex Procedure	R-T, ACC, FREQ
S2 Check COMPTMODE switch ...steer. Feedback: Tactile and Visual	Mission	02	H	Omit	Simple	ACC, FREQ

SEG 3: NAVIGATION TO TP	TASK: T3 CHECK FOR ACCURATE SYSTEM STEERING TO TP					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRICS
S1 Read system bearing and range to TP from DVRI bug and range displays. Feedback: Visual	Directive	05	H	Omit	Simple Procedure	TIME, ACC, FREQ
S2 Compare system bearing and range to TP with pre-planned or estimated bearing and range to TP. Feedback: Comparison Complete	Mission	05	H	Omit	Complex Procedure, Evaluation	TIME, ACC, SUBJ
S3 Go to T5 if system steering to TP is correct.	Directive	02	M	Sequence Commit	Simple Procedure	R-T, DEC
S4 Go to T4 if system steering to TP is not correct.	Directive	02	M	Sequence Commit	Simple Procedure	R-T, DEC

SEG 3: NAVIGATION TO TP	TASK: T4 TROUBLESHOOT SYSTEM STEERING IF REQUIRED					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRICS
S1 Determine TP Lat/Long from chart or IFR supplement. Feedback: Visual	Directive	08	M	Omit	Complex Procedure	TIME
S2 Compare system TP Lat/Long with actual TP Lat/Long. Feedback: Comparison Complete	Directive	12	M	Omit	Complex Procedure, Evaluation	TIME, ACC, SUBJ
(a) Throw DDU Data Switch ...on call. Feedback: Tactile	Sequential	02	M	Omit Sequence	Simple Procedure	T-S, FREQ
(b) Read system TGT N address (TP) from lower DDU Lat/Long displays. Feedback: Visual	Sequential	05	M	Omit	Simple Procedure	TIME
S3 Insert correct TP Lat/Long if required: for correct steering to TP. Feedback: Visual	Sequential	50	M	Omit Commit	Complex Procedure	TIME, T-S, ACC, FREQ

SEG 3: NAVIGATION TO TP | TASK: T4 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S4 Evaluate system present position and estimated position from chart: for accurate system steering. Feedback: Visual	Directive	67	H	Omit Commit	Complex Procedure, Evaluation	TIME, ACC, SUBJ
(a) Throw DDU Data Switch ....Pres Pos: for system present position. Feedback: Tactile	Sequential	02	M	Omit Sequence	Simple Procedure	T-S, FREQ
(b) Read system present position Lat/Long from lower DDU Lat/Long displays. Feedback: Visual	Sequential	05	M	Omit	Simple Procedure	TIME
(c) Compare system present position with estimated present position from chart. Feedback: Comparison Complete	Sequential	15	M	Omit	Complex Procedure, Evaluation	TIME, ACC, SUBJ

SEG 3: NAVIGATION TO TP | TASK: T4 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S4 (continued)						
(d) Insert correct present position if required. Feedback: Visual	Sequential	45	M	Omit	Complex Procedure	TIME, T-S, ACC, FREQ

SEG 3: NAVIGATION TO TP	TASK: T5 INFORM PILOT THAT SYSTEM STEERING IS TO THE TP					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 Check that pilot maintains safe flight and follows system steering Feedback: Vestibular Cues and Visual	Mission	20	H	Omit	Complex Procedure	R-T, ACC, FREQ, SIBJ

SEG 3: NAVIGATION TO TP	TASK: T6 INSERT DATA FOR NEXT REMAINING TPS IF REQUIRED					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 Throw COMPTMODE switch ....Enter. Feedback: Tactile and Visual	Mission	02	L	Omit Sequence	Simple Procedure	T-S, FREQ
S2 Depress previously utilized TGT N address key. Feedback: Tactile	Sequential	02	M	Omit	Complex Procedure	T-S, ACC, FREQ
S3 Depress POS action key. Feedback: Tactile	Sequential	02	L	Omit	Simple Procedure	T-S, FREQ
S4 Depress appropriate quantity keys in sequence for TP Lat/Long. Feedback: Visual	Sequential	09	M	Sequence	Complex Procedure	T-S, ACC, FREQ
S5 Depress ALT action key. Feedback: Tactile	Sequential	02	L	Omit Sequence	Simple Procedure	T-S, FREQ
S6 Depress appropriate quantity keys in sequence for TP altitude. Feedback: Visual	Sequential	09	M	Omit Sequence	Complex Procedure	T-S, ACC, FREQ

SEG 3: NAVIGATION TO TP | TASK: T6 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S7 Alert pilot to maintain present heading. Feedback: Auditory and Vestibular	Sequential	05	H	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S8 Throw COMPTMODE switch ....Steer. Feedback: Tactile and Visual	Sequential	02	L	Omit Sequence	Simple Procedure	T-S, FREQ
S9 Throw DDU DATA switch... on call. Feedback: Tactile and Visual	Directive	02	L	Omit	Simple Procedure	T-S, FREQ
S10 Check lower DDU Lat/Long displays for accurate data entry. Feedback: Visual	Sequential	10	M	Omit	Complex Procedure	TIME, ACC, FREQ, SUBJ
S11 Repeat S1 through S10 if required (inaccurate data entry performed). Feedback: Visual	Directive	45	M	Omit Commit Sequence	Complex Procedure	TIME, T-S, ACC, FREQ, DEC, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T6 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S12 Depress required <u>TGT N</u> Address key for current TP system steering Feedback: Tactile	Mission	03	H	Omit	Complex Procedure	T-S, ACC FREQ
S13 Check for accurate system steering to TP. Feedback: Evaluation Complete	Mission	10	H	Omit	Complex Procedure, Evaluation	TIME, T-S, ACC, FREQ, SUBJ
(a) Read system bearing and range to TP from DVRI bug and range displays. Feedback: Visual	Sequential	05	H	Omit	Simple Procedure	TIME, ACC, FREQ
(b) Compare system bearing and range to TP with pre-planned or estimated bearing and range to TP. Feedback: Comparison Complete	Sequential	05	H	Omit	Complex Procedure, Evaluation	TIME, ACC, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T6 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S14 Repeat S12 and S13 if required. Feedback: Visual	Inaccurate Steering	13	H	Commit	Complex Procedure, Evaluation	TIME, DEC
S15 Inform pilot that system steering is to the TP. Feedback: Auditory	Sequential	10	M	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
(a) Check that pilot maintains safe flight and follows system steering. Feedback: Vestibular Cues and Visual	Mission	05	H	Omit	Complex Procedure	ACC

SEG 3: NAVIGATION TO TP | TASK: T7 PERFORM SYSTEM NAVIGATION TASKS

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 Tune radar for optimum PPI display Feedback: Visual	Mission	27	H	Omit Commit	Complex Performance	TIME, T-S, FREQ, QUAL, SUBJ
(a) Adjust PPI Range control...CW (until TP at top of radar scope). Feedback: Visual	Sequential	06	H	Commit	Analysis	TIME, T-S, FREQ, SUBJ
(b) Adjust RCVR control ...CW/CCW (until return is enhanced). Feedback: Visual	Mission	08	H	Commit	Analysis	TIME, T-S, FREQ, QUAL, SUBJ
(c) Adjust STC Depth control...CW (until even return present). Feedback: Visual	Directive	08	M	Commit	Complex Procedure	TIME, T-S, FREQ, QUAL, SUBJ
(d) Adjust Scan Angle control...CW/CCW (until desired sweep width present). Feedback: Visual	Directive	05	M	Commit	Complex Procedure	TIME, T-S, FREQ, QUAL, SUBJ

SEG 3: NAVIGATION TO TP | TASK: T7 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S2 Monitor flight progress using radar significant terrain/cultural features as check points. Feedback: Accurate Navigational Position	Mission	15	H	Omit Commit	Complex Performance	TIME, T-S, ACC, FREQ, DEC, SUBJ
(a) Compare radar return image with preplanned chart and quick & dirty. Feedback: Accurate Navigation	Mission	15	H	Omit	Complex Performance	TIME, T-S, R-T, ACC, QUAL, SUBJ
(b) Compare system present position from chart if required. Feedback: Visual	Mission	20	H	Commit	Complex Procedure, Evaluation	TIME, T-S, ACC, FREQ, DEC
(c) Repeat (a) and (b) if required.	Mission	35	H	Commit Omit	Complex Performance	TIME, T-S, FREQ, DEC

SEG 3: NAVIGATION TO TP	TASK: T7 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S3 Inform pilot of system navigational accuracy. Feedback: Auditory	Directive	05	L	Omit	Complex Procedure	TIME, T-S, ACC, FREQ. SUBJ
S4 Position cursor intersection on TP if required. Feedback: Visual	Mission	23	H	Omit Commit	Complex Performance	TIME, T-S, R-T, ACC, FREQ, DEC, QUAL, SUBJ
(a) Place left hand on slew control stick. Feedback: Tactile	Radar Scope	01	H	Omit	Simple Procedure	T-S, R-T, FREQ
(b) Depress radar slew button and push slew control stick in direction of desired cursor intersection movement. Feedback: Visual	Sequential	05	H	Sequence	Complex Performance	TIME, T-S, ACC, FREQ, DEC, SUBJ
(c) Compare radar return image on cursor intersection with preplanned chart and quick & dirty TP. Feedback: Visual	Mission	15	H	Omit	Complex Performance	TIME, T-S, R-T, ACC, QUAL, SUBJ

SEG 3: NAVIGATION TO TP | TASK: T7 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S4 (continued)						
(d) Repeat (a) and (b) if required. Feedback: Accurate Positioning and Visual	Mission	06	H	Commit Omit	Complex Performance	TIME, T-S, ACC, FREQ, DEC, SUBJ
(e) Push Correct POS button on lower DDU panel. Feedback: Tactile	Sequential	02	H	Omit	Simple Procedure	ACC, FREQ
S5 Monitor navigational equipment for operating status or equipment condition. Feedback: Visual	Mission	18	H	Omit	Complex Performance	TIME, T-S, ACC, FREQ, SUBJ
(a) Check navigation (INS) control panel for INS failure indications. Feedback: Visual	Mission	02	H	Omit	Complex Procedure	R-T, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T7 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S5 (continued)						
(b) Check VTR panel for operating indications if required. Feedback: Visual	Directive	05	L	Omit	Complex Procedure	SUBJ
(c) Check for computer error light on DVRI panel and lower DDU panel. Feedback: Visual	Display	02	M	Omit	Simple Procedure	R-T, FREQ, SUBJ
(d) Check Attitude Ref switch...comp in. Feedback: Visual	Directive	02	M	Omit	Simple Procedure	FREQ, SUBJ
(e) Check magnetic variation from Mag Var display on lower DDU panel. Feedback: Visual	Directive	07	L	Omit	Complex Procedure	R-T, ACC, SUBJ
(f) Inform pilot of navigational equipment status. Feedback: Auditory.	Directive	05	L	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ

SEG 3: NAVIGATION TO TP | TASK: T7 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S6 *Monitor time on TP.* Feedback: Accurate Navigational Position	Mission	20	H	Omit	Complex Performance	*TIME*, T-S. ACC, FREQ, SUBJ
(a) Compare recorded actual leg or total time for previous TP with pre-planned leg or total time for previous TP. Feedback: Computation	Mission	15	H	Omit	Evaluation	TIME, ACC, FREQ, DEC, SUBJ
(b) Instruct pilot to adjust throttle controls so as to correct time on TP. Feedback: Accurate Mission Timing	Mission	15	H	Commit Omit Sequence	Complex Performance	R-T, ACC, FREQ, QUAL, SUBJ
(c) Inform pilot of time on TP results. Feedback: Auditory	Directive	05	L	Omit	Complex Procedure	TIME, T-S ACC, FREQ, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T7 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S7 Monitor system velocities Feedback: Small Cursor Drift	Mission	51	H	Omit	Complex Performance	TIME, T-S, R-T, ACC, FREQ, SUBJ
(a) Read system ground speed (GS) and wind. Feedback: Visual	Sequential	14	H	Omit	Complex Procedure	TIME, FREQ, SUBJ
(b) Read and record GS from doppler panel gnd speed display. Feedback: Visual	Sequential	05	M	Omit	Simple Procedure	TIME, FREQ, SUBJ
(c) Read and record indicated airspeed (IAS) from airspeed indicator on pilot's instrument panel. Feedback: Visual	Sequential	05	M	Omit	Simple Procedure	TIME, FREQ, SUBJ
(d) Read system true airspeed (TAS). Feedback: Visual	Sequential	07	M	Omit	Complex Procedure	TIME, FREQ, SUBJ

SEG 3: NAVIGATION TO TP | TASK: T7 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S7 (continued)						
(e) Evaluate system velocities using GS, wind, TAS, and IAS. Feedback: Evaluation Complete	Sequential	15	H	Omit	Evaluation	TIME, ACC, FREQ, SUBJ
(f) Troubleshoot system velocities if required. Feedback: Accurate System Velocities	Mission	60	H	Commit Omit	Analysis, Complex Performance	TIME, DEC, SUBJ
(g) Inform pilot of system velocity results. Feedback: Auditory	Directive	05	L	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S8 Monitor system heading. Feedback: Accurate Navigational Position	Mission	42	H	Omit	Complex Performance	TIME, T-S, R-T, ACC, FREQ, SUBJ

SEG 3: NAVIGATION TO TP — TASK: T7 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S8 (continued)						
(a) Read true heading from DVRI display bug Feedback: Visual	Sequential	05	H	Omit	Simple Procedure	TIME, ACC, SUBJ
(b) Read magnetic heading from wet compass. Feedback: Visual	Sequential	05	M	Omit	Simple Procedure	TIME, FREQ, SUBJ
(c) Read heading from horizontal situation indicator. Feedback: Visual	Sequential	05	M	Omit	Simple Procedure	TIME, FREQ, SUBJ
(d) Read magnetic variation from MAG VAR display on lower DDU panel. Feedback: Visual	Sequential	02	M	Omit	Simple Procedure	TIME, ACC, FREQ

SEG 3: NAVIGATION TO TP | TASK: T7 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S8 (continued)						
(e) Evaluate heading accuracy using pre-planned course, true heading, magnetic variation, and compass heading with "CDMVT" formula Feedback: Evaluation Complete	Mission	15	H	Omit Sequence	Complex Performance	TIME, T-S, ACC, DEC, SUBJ
(f) Adjust MA-1 compass needle deflection with pull to set control if required. Feedback: Visual	MA-1 Display	05	H	Omit	Complex Procedure	TIME, T-S, R-T, ACC, FREQ, DEC
(g) Troubleshoot system heading if required. Feedback: Accurate Headings	Mission	60	H	Omit Commit	Complex Performance	TIME, DEC, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T7 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S8 (continued)						
(h) Instruct pilot to adjust heading so as to maintain pre-planned course. Feedback: Visual	Mission	15	H	Omit Sequence	Complex Performance	TIME, T-S, ACC, FREQ, SUBJ
(i) Inform pilot of system heading results. Feedback: Auditory	Directive	05	L	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S9 Monitor system altitude. Feedback: Accurate system Altitudes.	Mission	25	L	Omit	Complex Performance	TIME, T-S, R-T, ACC, FREQ, SUBJ
(a) Read altitude (AGL) from radar altimeter. Feedback: Visual	Sequential	05	H	Omit	Simple Procedure	TIME, FREQ, SUBJ
(b) Read pressure altitude (MSL) from pressure altimeter. Feedback: Visual	Sequential	05	H	Omit	Simple Procedure	TIME, FREQ, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T7 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S9 (continued)						
(c) Read system altitude (MSL). Feedback: Visual	Sequential	05	L	Omit	Complex Procedure	TIME, FREQ, SUBJ
(d) Read terrain altitude for present position on chart. Feedback: Visual	Sequential	05	H	Omit	Complex Procedure	TIME, ACC, FREQ, SUBJ
(e) Evaluate system altitude accuracy using above sources. Feedback: Evaluation Complete	Sequential	15	L	Omit	Complex Performance	TIME, T-S ACC, FREQ, DEC, SUBJ
(f) Troubleshoot system altitude if required. Feedback: Accurate System Altitude	Mission	60	L	Commit Omit	Complex Performance	TIME, DEC, SUBJ
(g) Insert correct altitude if required. Feedback: Visual	Mission	25	L	Omit Commit	Complex Procedure	TIME, T-S ACC, FREQ, DEC, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T7 (continued)					
SUBTASK	**ACTION STIMULUS**	**TIME**	**CRIT**	**POTENTIAL ERROR**	**SKILLS REQUIRED**	**PERFORMANCE MEASURE METRIC**
S9 (continued)						
(h) Inform pilot of system altitude results. Feedback: Auditory	Directive	05	L	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S10 Monitor safety of flight instruments and equipment. Feedback: Safe Flight	Mission	37	H	Omit	Evaluation	TIME, T-S, R-T, ACC, FREQ, DEC, SUBJ
(a) Evaluate fuel status by comparing actual fuel remaining with pre-planned fuel remaining. Feedback: Evaluation Complete	Mission	15	H	Omit	Evaluation	TIME, ACC, FREQ, SUBJ
(b) Inform pilot of fuel status. Feedback: Auditory	Mission	05	H	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T7 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S10 (continued)						
(c) Check annunciator caution lights for emergency indications. Feedback: Visual	Mission	02	H	Omit	Simple Procedure	R-T, SUBJ
(d) Check fire warning lights for aircraft fire indications. Feedback: Visual	Mission	02	H	Omit	Simple Procedure	R-T, SUBJ
(e) Check accessible circuit breakers... in. Feedback: Tactile and Visual	Mission	08	H	Omit	Simple Procedure	R-T, SUBJ
(f) Inform pilot of safety of flight instruments and equipment results.	Mission	05	H	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ

SEG 3: NAVIGATION TO TP | TASK: T8 PERFORM APPROACH TO TP PROCEDURES

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 Check flight progress using radar significant terrain/cultural features as check points Feedback: Accurate Navigation	Mission	17	H	Omit Commit	Complex Performance	TIME, T-S, ACC, FREQ, DEC, SUBJ
(a) Compare radar return image on cursor intersection with pre-planned chart and quick & dirty. Feedback: Accurate Navigation	Mission	15	H	Omit	Complex Performance	TIME, T-S, R-T, ACC, QUAL, SUBJ
(b) Position cursor intersection on TP if required. Feedback: Visual	Mission	16	H	Omit Commit	Complex Performance	TIME, T-S, R-T, ACC, FREQ, DEL
(c) Repeat (a) and (b) if required. Feedback: Accurate Positioning and Visual	Mission	31	H	Omit	Complex Performance	TIME, T-S, FREQ, DEC, SUBJ

SEG 3: NAVIGATION TO TP — TASK: T8 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 (continued)						
(d) Push Correct Pos button on lower DDU panel. Feedback: Tactile	Sequential	02	H	Omit	Simple Procedure	ACC, FREQ
S2 Select ARE 30/60 display at approximately 17 miles from TP and navigation is accurate. Feedback: Tactile and Visual	Radar Scope	03	H	Omit Commit	Evaluation	R-T, ACC, FREQ, DEC, QUAL, SUBJ
S3 Observe ARE 30/60 expanded display at approximately 17 miles. Feedback: Visual	Directive	02	H	Omit Commit	Complex Procedure	SUBJ
S4 Tune radar for optimum ARE 30/60 display.	Mission	16	H	Omit	Complex Performance	TIME, T-S, R-T, FREQ, QUAL, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T8 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S4 (continued)						
(a) Rotate STC SLOPE control...CCW. Feedback: Tactile and Visual	Directive	02	M	Omit Sequence	Simple Procedure	TIME, T-S, FREQ
(b) Throw ANT PATT switch ...near (continue until brightest return present). Feedback: Visual	Directive	04	L	Omit Sequence	Complex Procedure	TIME, T-S, FREQ, SUBJ
(c) Adjust RCVR control control...CCW (until return is enhanced).	Directive	03	H	Omit Sequence	Synthesis	TIME, T-S, FREQ, QUAL
(d) Adjust SCAN ANGLE control...CW/CCW.	Directive	03	L	Omit Sequence	Simple Procedure	TIME, T-S, FREQ, QUAL
(e) Check AZ-RNG TRKG switch...off. Feedback: Visual	Directive	02	L	Omit Sequence	Simple Procedure	ACC, FREQ, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T8 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S4 (continued)						
(f) Check ELEV TRKG switch...off. Feedback: Visual	Directive	02	L	Omit	Simple Procedure	ACC, FREQ, SUBJ
(g) Adjust VIDEO/DIF Controls to enhance resolution if required.	Directive	10	L	Commit Omit	Synthesis	TIME, T-S, FREQ, QUAL
S5 Continue positioning cursor intersection on front leading edge center of turn point return. Feedback: Visual and Accurate Navigation.	Mission	05	H	Omit	Complex Performance	TIME, T-S, R-T, ACC, FREQ, QUAL, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T9 PERFORM VELOCITY CORRECT PROCEDURES					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS *REQUIRED*	PERFORMANCE MEASURE METRIC
S1 Go to T9.1 for automatic velocity correct (AZ-range lock-on or track-while-scan required).	Mission	02	H	Commit Omit	Evaluation	R-T, DEC
S2 Go to T9.2 for manual velocity correct.	Mission	02	H	Commit Omit	Evaluation	R-T, DEC
S3 FLIR tracking manual velocity correct (not current part of simulator capability).						

[illegible]: NAVIGATION TO TP	TASK: T9.1 PERFORM AUTOMATIC VELOCITY CORRECT					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 Position azimuth cursor to center of TP return. Feedback: Visual	Mission	03	H	Omit	Complex Performance	TIME, T-S, ACC, FREQ, QUAL
S2 Position range cursor to just leading edge of TP return. Feedback: Visual	Checklist	02	H	Omit	Complex Performance	TIME, T-S, ACC, FREQ, QUAL
S3 Throw AZ-RNG TRKG switch ...on. Feedback: Visual	Sequential	02	H	Omit	Simple Procedure	FREQ, SUBJ
S4 Check AZ-range indicator light on DVRI...on. Feedback: Visual	Sequential	02	H	Omit	Simple Procedure	FREQ, SUBJ
S5 Rotate velocity correct switch...memory point. Feedback: Visual and Tactile	Sequential	02	H	Omit	Simple Procedure	TIME, T-S, FREQ, SUBJ
S6 Check DDU data select switch...A. Feedback: Visual	Sequential	02	H	Omit	Simple Response	TIME, SUBJ

SEG 3: NAVIGATION TO TP | TASK: T9.1 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S7 Check flt data display 4 on upper DDU for some value greater than 000. Feedback: Visual	Sequential	03	H	Omit	Simple Procedure	TIME, SUBJ
S8 Rotate velocity correct switch...off save (before TP walk-down). Feedback: Visual, Tactile	Sequential	02	H	Omit	Simple Procedure	TIME, FREQ, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T9.1 PERFORM MANUAL VELOCITY CORRECT					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 Position azimuth cursor to center of TP return. Feedback: Visual	Mission	03	H	Omit	Complex Performance	TIME, T-S, ACC, FREQ, QUAL
S2 Position range cursor to just leading edge of TP return. Feedback: Visual	Checklist	02	H	Omit	Complex Performance	TIME, T-S, ACC, FREQ, QUAL
S3 Check <u>AZ-RNG TRKG</u> switch ...off. Feedback: Visual	Sequential	02	H	Omit	Simple Procedure	FREQ, SUBJ
S4 Rotate <u>Velocity Correct</u> switch...memory point. Feedback: Visual and Tactile	Sequential	02	H	Omit	Simple Procedure	TIME, T-S, FREQ, SUBJ
S5 Delay for 10 second minimum/128 second maximum to allow cursor drift. Feedback: Cockpit Clock	Sequential	10	H	Omit	Simple Procedure	TIME, T-S, R-T, FREQ, SUBJ

SEG 3: NAVIGATION TO TP — TASK: T9.2 (continued)

SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S6 Repeat positioning bearing and range cursors to just leading edge center of TP return. Feedback: Visual	Sequential	05	H	Omit	Complex Performance	TIME, T-S, ACC, FREQ, QUAL, SUBJ
S7 Monitor further cursor drift and repeat positioning if required Feedback: No Cursor Drift	Cursor Drift	15	H	Omit Commit	Complex Performance	TIME, T-S FREQ, DEC, SUBJ
S8 Check DDU Data Select switch...A. Feedback: Visual	Sequential	02	H	Omit	Simple Procedure	TIME, SUBJ
S9 Check Flt path display 4 on upper DDU panel for some value greater than 000. Feedback: Visual	Sequential	03	H	Omit	Simple Procedure	TIME, SUBJ
S10 Rotate Velocity Correct switch...off save (before TP walk-down). Feedback: Visual, tactile	Sequential	02	H	Omit	Simple Procedure	TIME, FREQ, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T10 INITIATE TURN AT TP					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S1 Alert pilot of next out-bound heading one minute prior to reaching TP. Feedback: Auditory	Directive	05	M	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S2 Set outbound heading on HSI. Feedback: Visual	Directive	05	L	Omit	Complex Procedure	TIME, T-S, ACC, FREQ
S3 Check AZ-RNG TRKG switch ...off. Feedback: Visual.	Directive	02	L	Omit	Simple Procedure	ACC, FREQ, SUBJ
S4 Monitor DVRI heading bug for movement to 180° relative position (TP passage) Feedback: Visual	Mission	48	H	Omit	Complex Procedure	R-T, FREQ
S5 Record leg time or total elapsed time at TP passage. Feedback: Visual	Mission	05	H	Omit	Complex Procedure	TIME, T-S, ACE, FREQ, SUBJ

SEG 3: NAVIGATION TO TP	TASK: T10 (continued)					
SUBTASK	ACTION STIMULUS	TIME	CRIT	POTENTIAL ERROR	SKILLS REQUIRED	PERFORMANCE MEASURE METRIC
S6 Activate cockpit clock at TP passage if required (leg time only). Feedback: Tactile and Visual	Directive	02	H	Omit	Complex Procedure	R-T, FREQ
S7 Inform pilot of TP passage and outbound heading. Feedback: Auditory	Sequential	05	M	Omit	Complex Procedure	TIME, T-S, ACC, FREQ, SUBJ
S8 Check for pilot turning to new heading. Feedback: Vestibular and Visual	Sequential	30	H	Omit	Simple Procedure	ACC
For subsequent navigation legs, return to Segment 3, Task 2.						

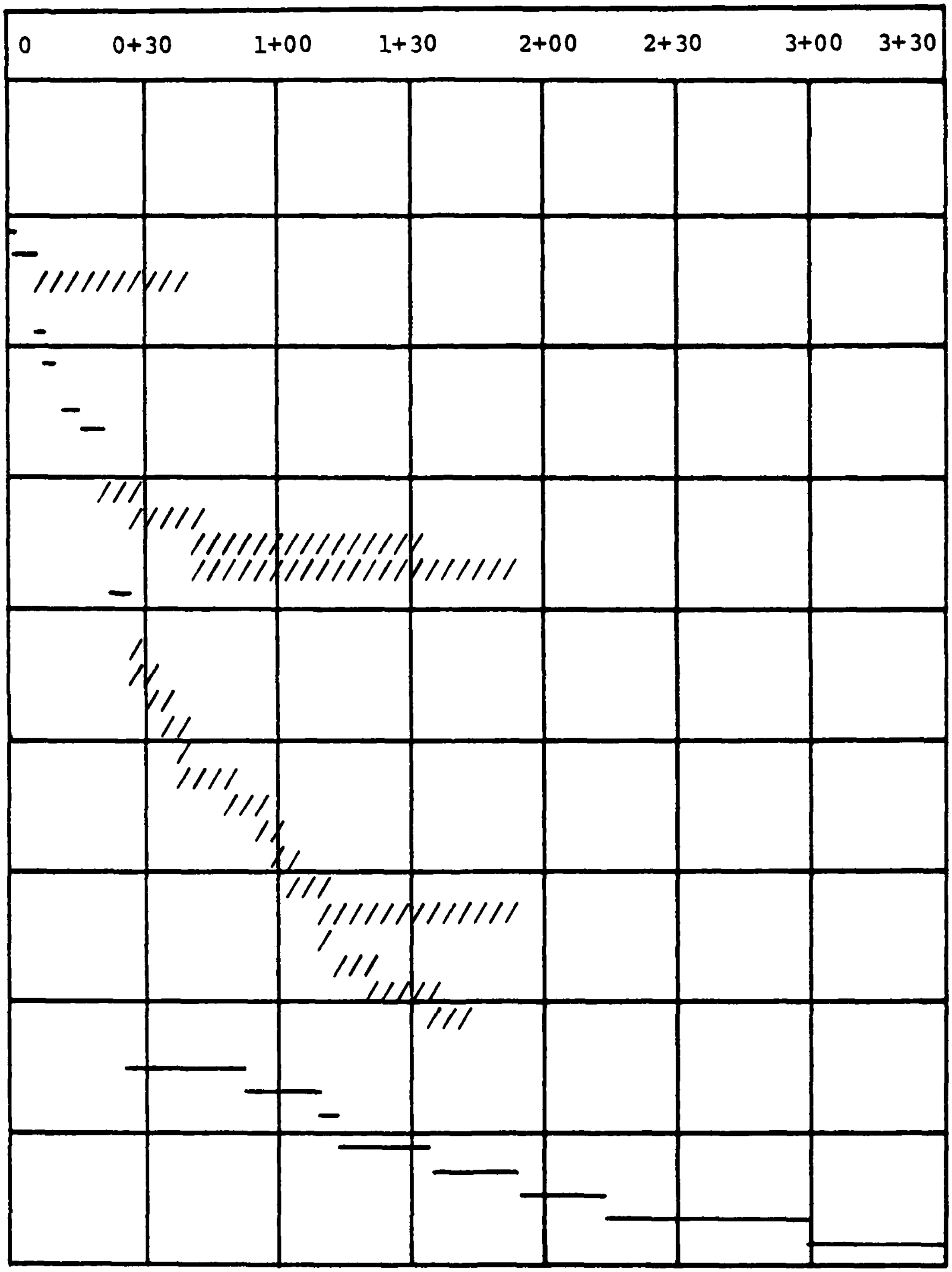
APPENDIX D

RADAR NAVIGATION MISSION TIME LINE ANALYSIS

The Mission Time Line Analysis (MTLA) relates the sequence of tasks to be performed by the operator to a real time basis, and can be used to identify critical tasks within a maneuver that are important for performance measurement [Matheny, et al., 1970]. Using the segment three portion of the A-6E TRAM radar navigation task listing (Appendix A), each task/subtask was listed along the vertical axis of the time line. The estimated time to perform each task and subtask was then extracted from the task analysis (Appendix C) and plotted along the horizontal axis, which represents in this example a seven-minute radar navigation TP-to-TP "leg." Time is coded as: (1) dark if the task must be executed for maneuver success or if the task requires complete operator attention, or (2) shaded if the task is one of monitoring or troubleshooting and can be performed simultaneously with other tasks.

The MTLA is a large graph but is presented here as two task pages (T1 to T7, S8; and T7, S9 to T10) each followed by two time pages (0 to 3+30, and 3+30 to 7+00). By removing and appropriately arranging the six pages, the full MTLA graph will result.

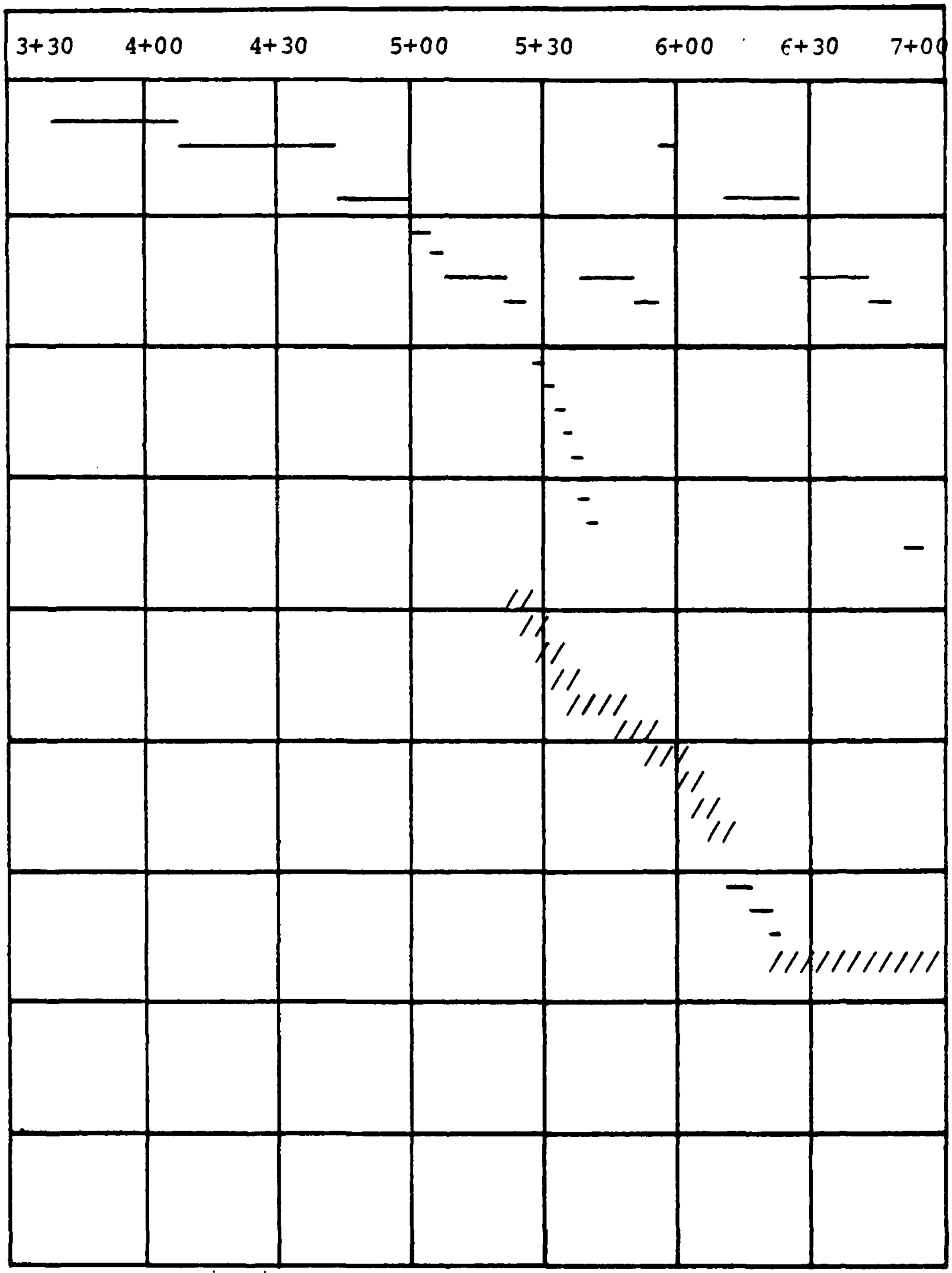
SEGMENT 3: NAVIGATION TO TURN POINT (TP)	-0+30	0
T1 INITIATE TURN AT IP S1 ALERT PILOT OF OUTBOUND HEADING S2 SET HEADING ON HSI S3 CHECK AZ-RNG TRKG SWITCH S4 MONITOR DVRI FOR IP PASSAGE	//////	////////
S5 ACTIVATE COCKPIT CLOCK S6 INFORM PILOT OF IP PASSAGE S7 CHECK FOR PILOT TURNING T2 ACTIVATE STEERING TO TP S1 DEPRESS TGT N ADDRESS KEY		
S2 CHECK COMPTMODE SWITCH T3 CHECK FOR ACCURATE STEERING S1 READ SYSTEM BEARING AND RANGE S2 COMPARE BEARING AND RANGES T4 TROUBLESHOOT STEERING IF REQUIRED		
S1 DETERMINE ACTUAL TP LAT/LONG S2 COMPARE ACTUAL/SYSTEM TP S3 INSERT CORRECT TP LAT/LONG S4 EVALUATE SYSTEM PRESENT POSITION T5 INFORM PILOT OF STEERING TO TP		
T6 INSERT DATA FOR NEXT TP(s) S1 THROW COMPTMODE SWITCH S2 DEPRESS OLD TGT N ADDRESS KEY S3 DEPRESS POS ACTION KEY S4 DEPRESS QUANTITY KEYS		
S5 DEPRESS ALT ACTION KEY S6 DEPRESS QUANTITY KEYS S7 ALERT PILOT TO MAINTAIN HEADING S8 THROW COMPTMODE SWITCH S9 THROW DDU DATA SWITCH		
S10 CHECK FOR ACCURATE DATA ENTRY S11 REPEAT INSERT IF REQUIRED S12 DEPRESS TGT N ADDRESS KEY FOR TP S13 CHECK FOR ACCURATE STEERING S14 REPEAT S12/S13 IF REQUIRED		
S15 INFORM PILOT OF STEERING TO TP T7 PERFORM SYSTEM NAVIGATION TASKS S1 TUNE RADAR FOR OPTIMUM DISPLAY S2 MONITOR FLIGHT PROGRESS S3 INFORM PILOT OF NAV ACCURACY		
S4 POSITION CURSORS ON TP S5 MONITOR NAVIGATION EQUIPMENT S6 MONITOR TIME ON TP S7 MONITOR SYSTEM VELOCITIES S8 MONITOR SYSTEM HEADING		


0
0+30
1+00
1+30
2+00
2+30
3+00
3+30

3+30 4+00 4+30 5+00 5+30 6+00 6+30 7+00

SEGMENT 3: (continued)	-0+30	0
T7 (continued)		
S9 MONITOR SYSTEM ALTITUDE		
S10 MONITOR FLIGHT SAFETY INSTR.		
T8 PERFORM APPROACH TO TP PROCEDURES		
S1 CHECK FLIGHT PROGRESS		
S2 SELECT ARE 30/60 DISPLAY		
S3 OBSERVE ARE 30/60 DISPLAY		
S4 TUNE RADAR FOR OPTIMUM DISPLAY		
S5 CONTINUE CURSOR POSITIONING		
T9.1 PERFORM AUTOMATIC VELOCITY CORRECT		
S1 POSITION AZIMUTH CURSOR		
S2 POSITION RANGE CURSOR		
S3 THROW AZ-RNG TRKG SWITCH		
S4 CHECK FOR AZ-RNG LOCK-ON		
S5 ROTATE VELOCITY CORRECT SWITCH		
S6 CHECK DDU DATA SELECT SWITCH		
S7 CHECK A-4 DISPLAY		
S8 ROTATE VELOCITY CORRECT SWITCH		
T9.1 PERFORM MANUAL VELOCITY CORRECT		
S1 POSITION AZIMUTH CURSOR		
S2 POSITION RANGE CURSOR		
S3 CHECK AZ-RNG TRKG SWITCH		
S4 ROTATE VELOCITY CORRECT SWITCH		
S5 DELAY FOR 10-SEC MINIMUM		
S6 REPEAT CURSOR POSITIONING		
S7 MONITOR FURTHER CURSOR DRIFT		
S8 CHECK DDU DATA SELECT SWITCH		
S9 CHECK A-4 DISPLAY		
S10 ROTATE VELOCITY CORRECT SWITCH		
T10 INITIATE TURN AT TP		
S1 ALERT PILOT OF OUTBOUND HEADING		
S2 SET HEADING ON HSI		
S3 CHECK AZ-RNG TRKG SWITCH		
S4 MONITOR DVRI FOR TP PASSAGE		
S5 RECORD TIME OF TP PASSAGE		
S6 ACTIVATE COCKPIT CLOCK		
S7 INFORM PILOT OF TP PASSAGE		
RETURN TO SEGMENT 3, TASK 2		

0	0+30	1+00	1+30	2+00	2+30	3+00	3+30


3+30
4+00
4+30
5+00
5+30
6+00
6+30
7+00

APPENDIX E

SEQUENTIAL SAMPLING DECISION MODEL

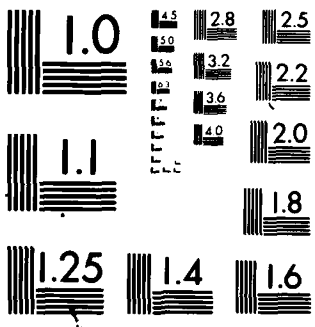
This appendix presents the sequential sampling decision model and its parameters in sufficient detail for the reader unfamiliar with the background theory of the model. Much of the material is excerpted from Rankin and McDaniel [1980] in a method proposal for achieving improvements in the precision of determining FRS student aviator proficiency using a Computer Aided Training Evaluation and Scheduling (CATES) system. CATES provides a computer managed, prescriptive training program based on individual student performance, and could be utilized for the evaluation portion of the model to measure B/N performance by either simulator software incorporation or by desk-top minicomputers.

AD-A098 776   NAVAL POSTGRADUATE SCHOOL MONTEREY CA   F/G 5/9
A MODEL TO MEASURE BOMBARDIER/NAVIGATOR PERFORMANCE DURING RADA--ETC(U)
MAR 81 T R MIXON

UNCLASSIFIED   NL

4 of 4

END
DATE
FILMED
8-81
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

I. CATES DECISION MODEL

One sequential method that may be used as a means for making statistical decisions with a minimum sample was introduced by Wald [1947]. Probability ratio tests and corresponding sequential procedures were developed for several statistical distributions. One of the tests, the binomial probability ratio test, was formulated in a context of a sampling procedure to determine whether a collection of a manufactured product should be rejected because the proportion of defectives is too high or should be accepted because the proportion of the defectives is below an acceptable level. The sequential testing procedure also provides for a postponement of decisions concerning acceptance or rejection. This deferred decision is based on prescribed values of alpha (α) and beta (β). Alpha (α) limits errors of declaring something "True" when is is "False" (Type I error). Beta (β) limits errors of declaring something "False" when it is "True" (Type II error).

In an industrial quality control setting, the inspector needs a chart similar to Figure E1 to perform a sequential test to determine if a manufacturing process has turned out a lot with too many defective items or whether the proportion of defects is acceptable. As each item is observed, the inspector plots a point on the chart one unit to the right if it is not defective, one unit to the right and one unit up if the item is defective. If the plotted line crosses the upper parallel line, the inspector will reject the production lot.

If the plotted line crosses the lower parallel line, the lot will be accepted. If the plotted line remains between the two parallel lines of the sequential decision chart, another sample item will be drawn and observed/tested.

The CATES decision model focuses on proportions of proficient trials (analogous to nondefectives or correct responses) whereas, in previous applications, proportions of defectives or incorrect responses were the items of interest. This approach does not alter the logic of the sequential sampling procedure or the decision model. It does enhance the "meaningfulness" of the procedure in decisions concerning proficiency because the ultimate goal is to determine "proficiency" rather than "nonproficiency." It should be noted that in the industrial quality control setting, sampling occurs after the manufacturing process. In educational and training applications, sequential sampling occurred <u>after</u> the learning period. In the CATES System, the sequential sampling occurs <u>during</u> the learning period and eventually terminates it.

The CATES decision model can be described as consisting of decision boundaries. Referring to Figure E1, the parallel lines represent those decision boundaries. Crossing the upper line, or boundary, results in a decision to "Reject Lot"; crossing the lower line, or boundary, results in a decision to "Accept Lot." In the CATES system, these decision boundaries translate to "Proficient" and "Not Proficient." Calculations of the decision boundaries require four parameters. These four parameters are:

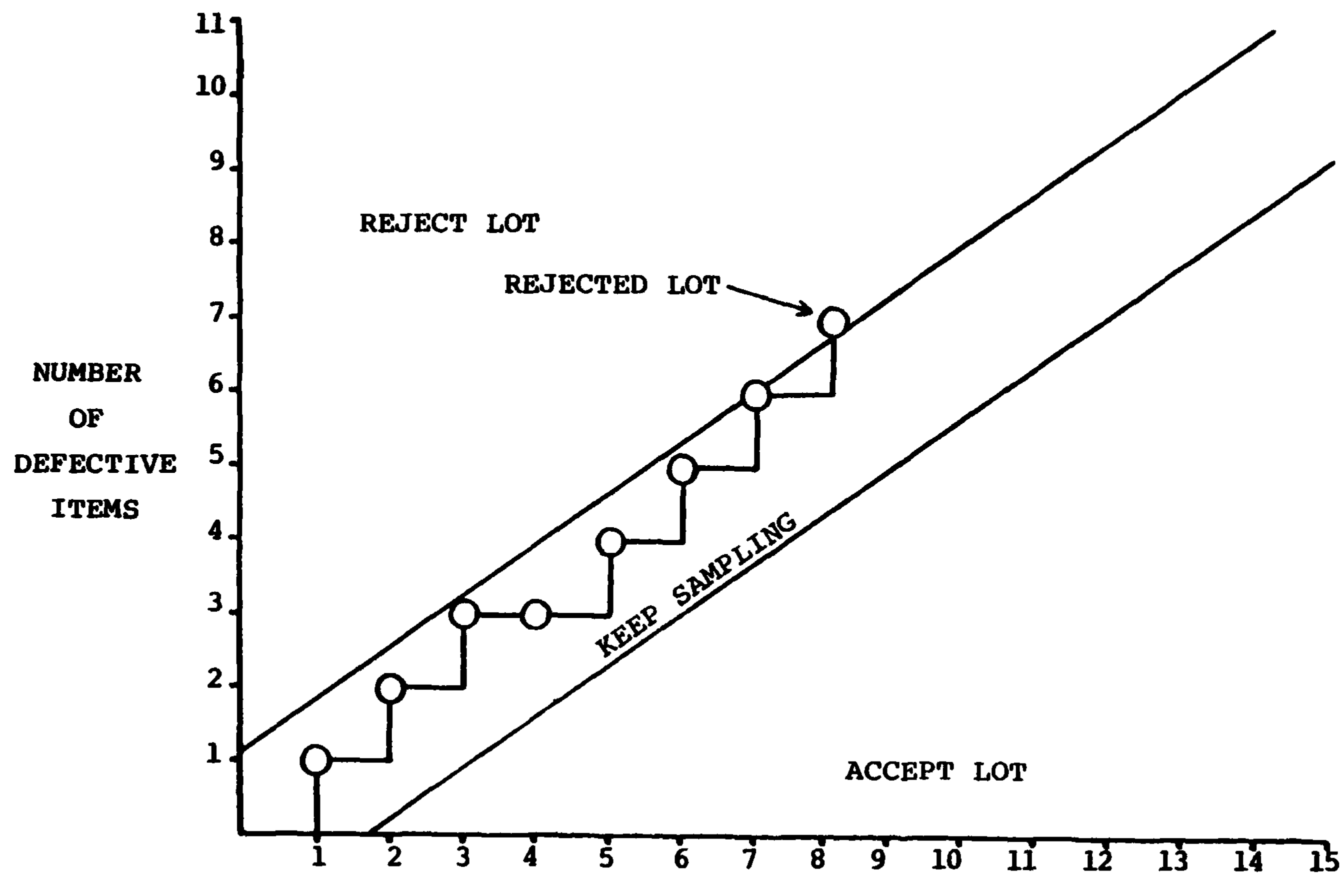


Figure E1. Hypothetical Sequential Sampling Chart.

P_1 Lowest acceptable proportion of proficient trials (P) required to pass the NATOPS flight evaluation with a grade of "Qualified." Passage of the NATOPS flight evaluation is required to be considered a trained aviator in an operational (fleet) squadron.

P_2 Acceptable proportion of proficient trials (P) that represent desirable performance on the NATOPS flight evaluation.

Alpha (α) The probability of making a TYPE I decision error (deciding a student is proficient when in fact he is not proficient).

Beta (β) The probability of making a TYPE II decision error (deciding a student is not proficient when in fact he is proficient.

Parameter setting is a crucial element in the development of the sequential sampling decision model. Kalisch [1980] outlines three methods for selecting proficient/not proficient performance (q_0/q_1 values) as:

Method 1--External Criterion. Individuals are classified as masters, non-masters, or unknown on the basis of performance on criteria directly related to the instructional objectives. These criteria can be in terms of demonstrated levels of proficiency either on the job or in a training environment, The mean proportion of items answered correctly by the masters on an objective would provide an estimate for q_0. Similarly, q_1 would be the proportion correct for the non-masters.

Method 2--Rationalization. Experts in the subject area who understand the relation of the training objectives to the end result; e.g., on-the-job performance, select the q_0 and q_1 values to reflect their estimation of the necessary levels of performance. This method is probably the closest to that now used by the Air Force. The procedure may provide somewhat easier decision making since specifying two values creates an indecision zone--neither mastery nor non-mastery. This indecision zone indicates that performance is at a level which may not be mastery but is not sufficiently poor to be considered at a non-mastery level.

Method 3--Representative Sample. The scores of prior trainees, who demonstrate the entire range from extremely poor to exemplary performance on objectives, are used to estimate q_0 and q_1. The proportion correct for the entire sample is used to obtain an initial cutting score C. Scores are separated into two categories: (a) those scores greater than or equal to C and (b) those scores less than C. For each category, the mean proportion correct score is computed. The mean for the first category equals q_0; the mean for the second category equals q_1.

The selection of alpha (α) and beta (β) should be based on the criticality of accurate proficiency decisions. Small values of alpha (α) and beta (β) require additional task trials to make decisions with greater confidence. Factors that are important in selecting values for alpha (α) and beta (β) are outlined below:

(1) Alpha (α) values

(a) Safety--potential harm to the trainee or to others due to the trainee's actual non-mastery of the task.

(b) Prerequisite in Instruction--potential problems in future instruction, especially if the task is prerequisite to other tasks.

(c) Time/Cost--potential loss or destruction of equipment either in training or upon fleet assignment.

(d) Trainee's View of the Training--potential negative view by trainee when classified as proficient although the trainee lacks confidence in that decision. Also, after fleet assignment if previous training has not prepared him sufficiently the trainee may also have a negative view of the training program.

(2) Beta (β) values

(a) Instruction--requirement for additional training resources (personnel and materials) for unnecessary training in case of misclassification as not proficient.

(b) Trainee Attitudes--the attitude of trainees when tasks have been mastered yet training continues; trainee frustration; corresponding impact on performance in the remainder of the training program and fleet assignment.

(c) Cost/Time--the additional cost and time required for additional training that is not really needed.

After the model parameters have been selected, calculation of the decision boundaries may be accomplished using the Wald Binomial Probability Ratio Test. A formal mathematical discussion of this test follows.

II. WALD BINOMIAL PROBABILITY RATIO TEST

The Wald binomial probability ratio test was developed by Wald [1947] as a means of making statistical decisions using as limited a sample as possible. The procedure involves the consideration of two hypotheses:

$$H_0: \; P \leq P_1$$

and $H_1: \; P \geq P_2$ where

P is the proportion of nondefectives in the collection under consideration, P_1 is the minimum proportion of nondefectives at or below which the collection is rejected, and P_2 is the desired proportion of nondefectives, at or above which the collection is accepted. Since a simple hypothesis is being tested against a simple alternative, the basis for deciding between H_0 and H_1 may be tested using the likelihood ratio:

$$\frac{P_{2n}}{P_{1n}} = \frac{(P_2)^{dn}\ (1 - P_2)^{n-dn}}{(P_1)^{dn}\ (1 - P_1)^{n-dn}}$$

Where: P_1 = Minimum proportion of nondefectives at or below which the collection is rejected.

P_2 = Desirable proportion of nondefectives at or above which the collection is accepted.

n = Total items in collection.

dn = Total nondefectives in collection.

The sequential testing procedure provides for a postponement region based on prescribed values of alpha (α) and beta (β) that approximate the two types of errors found in the statistical decision process. To test the hypothesis H_0: $P = P_1$, calculate the likelihood ratio and proceed as follows:

(1) If $\frac{P_{2n}}{P_{1n}} \leq \frac{\beta}{1-\alpha}$, accept H_0

(2) If $\frac{P_{2n}}{P_{1n}} \geq \frac{1-\beta}{\alpha}$, accept H_1

(3) If $\frac{\beta}{1-\alpha} < \frac{P_{2n}}{P_{1n}} < \frac{1-\beta}{\alpha}$, take an additional observation.

These three decisions relate well to the task proficiency problem. We may use the following rules:

(1) Accept the hypothesis that the grade of P is accumulated in lower proportions than acceptable performance would indicate.

(2) Reject the hypothesis that the grade of P is accumulated in lower proportions than acceptable performance would indicate. By rejecting this hypothesis, an alternative hypothesis is accepted that the grade of P is accumulated in proportions equal to or greater than desired performance.

(3) Continue training by taking an additional trial(s); a decision cannot be made with specified confidence.

The following equations are used to calculate the decision regions of the sequential sampling decision model.

$$dn \leq \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{P_2}{P_1} + \log \frac{1-P_1}{1-P_2}} + n \frac{\log \frac{1-P_1}{1-P_2}}{\log \frac{P_2}{P_1} + \log \frac{1-P_1}{1-P_2}}$$

$$dn \geq \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{P_2}{P_1} + \log \frac{1-P_1}{1-P_2}} + n \frac{\log \frac{1-P_1}{1-P_2}}{\log \frac{P_2}{P_1} + \log \frac{1-P_1}{1-P_2}}$$

Where: dn = Accumulation of trials graded as "P" in the sequence.

n - Total trials presented in the sequence.

P_1 = Lowest acceptable proportion of proficient trials (P) required to pass the NATOPS flight evaluation with a grade of "Qualified."

P_2 = Proportion of proficient trials (P) that represent desirable performance on the NATOPS flight evaluation.

Alpha (α) = The probability of making a type I error (deciding a student is proficient when in fact he is not proficient.

Beta (β) = The probability of making a type II error (deciding a student is not proficient when in fact he is proficient).

The first term of the two equations will determine the intercepts of the two linear equations. The width between these intercepts is determined largely by values selected for alpha (α) and beta (β). The width between the intercepts translates into a region of uncertainty; thus, as lower values of alpha (α) and beta (β) are selected this region of uncertainty increases.

The second term of the equations determines the slopes of the linear equation. Since the second term is the same for both equations, the result will be slopes with parallel lines. Values of P_1 and P_2 as well as differences between P_1 and P_2 affect the slope of the lines. This is easily translated into task difficulty. As P_2 values increase, indicating easier tasks, the slope becomes more steep. This in turn results in fewer trials required in the sample to reach a decision.

As differences in P_1 and P_2 increase, the slope also becomes steeper and the uncertainty region decreases. This is consonant with rational decision making. When the difference between the lower level of proficiency and upper level of proficiency is great, it is easier to determine at which proficiency level the pilot trainee is performing. The concept of differences in P_1 and P_2 is analogous to the concept of effect size in statistically testing the difference between the means of two groups. In such statistical testing, when alpha (α) and beta (β) remain constant, the number of observations required to detect a significant difference may be reduced as the anticipated effect size increases [Kalisch, 1980].

INITIAL DISTRIBUTION LIST

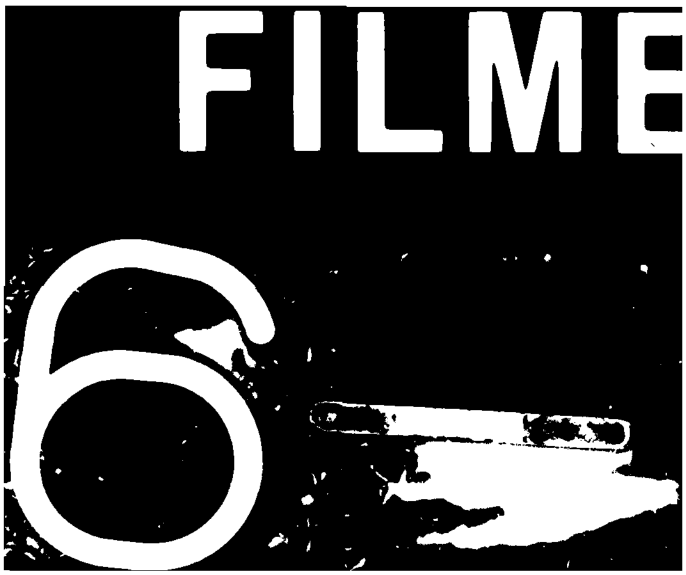
		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
4.	Asst Professor D. E. Neil, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	2
5.	LT Ted R. Mixon 461 Lilac Rd Casselberry, Florida 32707	5
6.	Mr. Don Vreuls Vreuls Research Corporation 68 Long Court, Suite E Thousand Oaks, California 91360	1
7.	Dr. Jack B. Shelnutt Seville Research Corporation 400 Plaza Building Pensacola, Florida 32505	1
8.	A. F. Smode, Director TAEG Training Analysis and Evaluation Group Orlando, Florida 32813	1
9.	LT Gerald S. Stoffer Code N-712 Naval Training Equipment Center Orlando, Florida 32813	1
10.	Don Norman Code N-71 Naval Training Equipment Center Orlando, Florida 32813	1

No. Copies

11.	COMNAVAIRPAC, Commander Naval Air Force US Pacific Fleet Box 1210, Naval Air Station, North Island San Diego, California 92135	1
12.	COMNAVAIRLANT, Commander Naval Air Force US Atlantic Fleet Norfolk, Virginia 23511	1
13.	COMTACWINGSLANT, Commander Tactical Wings Atlantic Naval Air Station, Oceana Virginia Beach, Virginia 23460	1
14.	COMMATVAQWINGPAC, Commander Medium Attack Tactical Electronic Warfare Wing Pacific Naval Air Station, Whidbey Island Oak Harbor, Washington 98278	2
15.	MATWING, Medium Attack Wing 1 Naval Air Station, Odeana Virginia Beach, Virginia 23460	2
16.	Commanding Officer, Attack Squadron 42 (VA-42) Naval Air Station, Oceana Virginia Beach, Virginia 23460	2
17.	Commanding Officer, Attack Squadron 128 (VA-128) Naval Air Station, Whidbey Island Oak Harbor, Washington 98278	2
18.	CDR W. Moroney, Code 55Mp Department of Operations Research Naval Postgraduate School Monterey, California 93940	6
19.	NAVAIRSYSCOM Headquarters - AIR 5313 Navy Department Washington, DC 20361	1
20.	NAVAIRSYSCOM Headquarters - AIR 340 F Navy Department Washington, DC 20361	1
21.	NAVAIRSYSCOM Headquarters - AIR 4135 A Navy Department Washington, DC 20361	1
22.	Leonard Ryan Naval Training Equipment Center Orlando, Florida 32813	1

		No. Copies
23.	Dr. Sam Campbell Code A11-04 Grumman Aerospace Corporation Bethpage, New York 11714	1
24.	Professor Glenn Lindsay, Code 55Ls Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
25.	Dr. Tom Killion Air Force Human Resources Laboratory Code OT(UDRI), Bldg 558 Williams AFB, Arizona 85224	1
26.	Ira Goldstein Code N-71 Naval Training Equipment Center Orlando, Florida 32813	1
27.	LT Alexander Bory Human Factors Engineering Branch Pacific Missile Test Center Point Mugu, California 93042	1
28.	CDR Paul R. Chatelier Office of Director Defense Research and Development 3 D 129 Pentagon Washington, DC 20350	1
29.	LCDR Wade R. Helm Human Factors Division Naval Air Development Center Warminster, Pennsylvania 18074	1
30.	CDR Robert S. Kennedy Chief, Human Performance Science Division Naval Biodynamic Laboratory Box 29407, Michoud Station New Orleans, Louisiana 70189	1
31.	CDR Norman E. Lane Human Factors Engineering Division Crew Systems Department NADC Warminster, Pennsylvania 18074	1

	No. Copies
32. LT Michael G. Lilienthal Code 05 Human Performance Sciences Department Naval Aerospace Medical Research Laboratory NAS Pensacola, Florida 32508	1
33. CDR Joseph F. Funaro Code N-82 Naval Training Equipment Center Orlando, Florida 32813	1
34. CAPT Thomas J. Gallagher Naval Air Development Center Warminster, Pennsylvania 18074	1
35. LCDR Richard H. Shannon Naval Biodynamic Laboratory Box 29407, Michoud Station New Orleans, Louisiana 70189	1
36. CDR Patrick M. Curran Head, Aerospace Psychology Branch (3C13) Bureau of Medicine and Surgery Navy Department Washington, DC 20372	1
37. Stan Collyer Human Factors Laboratory Naval Training Equipment Center Orlando, Florida 32813	1
38. Larry R. Beideman McDonnell Douglas Corporation Box 516 Saint Louis, Missouri 63166	1
39. William C. McDaniel Training Analysis and Evaluation Group Orlando, Florida 32813	1


FILME
6